Psychometrika

VOLUME X—1945
JANUARY-DECEMBER

PUBLISHED QUARTERLY
By THE PSYCHOMETRIC SOCIETY
AT 23 WEST COLORADO AVENUE
COLORADO SPRINGS, COLORADO

Scien

P

VO
NU
M
1

Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE

THE PSYCHOMETRIC SOCIETY · ORGANIZED IN 1935

VOLUME 10
NUMBER 1
MARCH
1945

PSYCHOMETRIKA, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

MARCH 1945, VOLUME 10, NUMBER 1

Printed for the Psychometric Society at 23 West Colorado Avenue, Colorado Springs, Colorado. Entered as second class matter, September 17, 1940, at the Post Office of Colorado Springs, Colorado, under the act of March 3, 1879. Editorial Office, College Entrance Examination Board, Princeton, New Jersey.

Subscription Price:

To non-members, the subscription price is $5.00 per volume of four issues. Members of the Psychometric Society pay annual dues of $5.00, of which $4.50 is in payment of a subscription to *Psychometrika.* Student members of the Psychometric Society pay annual dues of $3.00, of which $2.70 is in payment for the journal.

The subscription price to libraries and other institutions is $10.00 per year; this price includes one extra copy of each issue.

Applications for membership and student membership in the Psychometric Society should be sent to

THELMA GWINN THURSTONE
Chairman of the Membership Committee
University of Chicago
Chicago, Illinois

Membership dues are payable to

IRVING D. LORGE
Treasurer of the Psychometric Society
425 West 123rd Street
New York, New York

Individual and library subscriptions are payable to, and change of address should be sent to

IRVING D. LORGE
Treasurer of the Psychometric Corporation
425 West 123rd Street
New York, New York

Articles on the following subjects are published in *Psychometrika*:

(1) the development of quantitative rationale for the solution of psychological problems;
(2) general theoretical articles on quantitative methodology in the social and biological sciences;
(3) new mathematical and statistical techniques for the evaluation of psychological data;
(4) aids in the application of statistical techniques, such as monographs, tables, work-sheet layouts, forms, and apparatus;
(5) critiques or reviews of significant studies involving the use of quantitative techniques.

(*Continued on the back inside cover page*)

Psychometrika

CONTENTS

VOLUME TEN MARCH 1945 NUMBER ONE

PSYCHOMETRIKA—VOL. 10, NO. 1
MARCH, 1945

THE EFFECT OF DIFFICULTY AND CHANCE SUCCESS ON CORRELATIONS BETWEEN ITEMS OR BETWEEN TESTS

JOHN B. CARROLL
ENS., H(S), U. S. N. R.*

A study is made of the extent to which correlations between items and between tests are affected by the difficulties of the items involved and by chance success through guessing. The Pearsonian product-moment coefficient does not necessarily give a correct indication of the relation between items or sets of items, since it tends to decrease as the items or tests become less similar in difficulty. It is suggested that the tetrachoric correlation coefficient can properly be used for estimating the correlation between the continua underlying items or sets of items even though they differ in difficulty, and a method for correcting a 2 × 2 table for the effect of chance is proposed.

The correlation coefficient has frequently been used as an indication of the extent to which two items or two tests measure the same ability. It is the purpose of this paper to show that correlations between items and between tests are affected by the difficulties of the items involved, and that to the degree that the items or tests are dissimilar in difficulty, conventional correlational statistics tend not to give a correct indication of the true overlap of ability. This demonstration is made through the analysis of a theoretical limiting case, that is, one in which a set of items measure a single ability, but it can also be shown to apply whenever items have any factorial overlap. ? a priori?!

Assume that we have a set of n perfectly reliable items of varying difficulty which all measure a single ability and only a single ability. Difficulty is measured by the proportion (k) of individuals failing each item. It is assumed that in order to pass an item the individual must have true mastery of the task involved; that is, the probability of chance success, c, is assumed to be zero. It is further assumed that each item has been presented to all individuals under constant conditions.† The ability measured is of such a nature that success at any level of difficulty implies success at all lower levels of

* The opinions expressed in this article are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large. The writer is indebted to Lt. C. L. Vaughn, H(S) USNR, for critical comments on this paper.

† This assumption precludes the analysis of a set of items in a time-limit test where the subjects are exposed to varying numbers of items.

difficulty, and failure at any level implies failure at all higher levels of difficulty.* Thus, it will be found that all individuals who pass an item at any given level of difficulty will pass all items of less difficulty, and that all individuals who fail an item at any given level of difficulty will fail all items of greater difficulty.†

Statistics of a Set of n *Items Varying in Difficulty*

For convenience in later formulations, all statistics will be developed in terms of failure scores, *i.e.*, where failure on an item is scored as 1, passing as 0.

Let E_i be the failure score of an individual on item i in a test of n items. Let $E = \sum_{i=1}^{n} E_i$; that is, E represents the total score of an individual on the test of n items. N will signify the number of individuals in the population. If we let k_i represent the proportion of individuals failing item i, we find

$$k_i = \frac{\sum^{N} E_i}{N}, \tag{1}$$

and that the mean score on the test of n items is written as

$$\bar{E} = \frac{1}{N} \sum^{N} \sum^{n} E_i = \sum_{i=1}^{n} k_i . \tag{2}$$

The sum of squared test scores is found as

$$\begin{aligned} \sum^{N} E^2 &= \sum^{N} [E_1 + E_2 + E_3 + \cdots + E_i + E_j + \cdots + E_n]^2 \\ &= \sum E_1^2 + \sum E_2^2 + \sum E_3^2 + \cdots + \sum E_i^2 + \sum E_j^2 \\ &\quad + \cdots + \sum E_n^2 + 2 \sum_{i \neq j} E_i E_j , \end{aligned} \tag{3}$$

where i refers to the easier of a pair of items, j to the harder of the pair. Subscripts $1, 2, 3, \cdots, i, j, \cdots, n$ refer to items. Since all scores are either 1 or 0, $\sum E_i^2 = \sum E_i$; therefore, the sum of the squared expressions in (3) can be written merely as $N \sum k_i$, by (2). The expression $2\sum E_i E_j$ represents the cross-products of scores on pairs of items. Only those failing both items of a pair will contribute a term other than zero to the sum $\sum E_i E_j$. In accordance with our

* Most factors of ability, but not necessarily all, are probably of this nature.

† If one moves out of the context of these assumptions, however, the fact that any given pair of items is characterized by such a relation does not guarantee that the items are factorially homogeneous.

assumptions, the number of individuals who fail both items of a pair is given immediately by the number who fail the easier item, since none of these pass the harder item. If all n items are ranked in difficulty,

$$\sum E_i E_j = N \sum n_a k_i , \tag{4}$$

where $n_a =$ the number of items ranked above item i in order of increasing difficulty. The expression n_a is intended to include all items ranked above i even when equal in difficulty to item i. Then,

$$\sum E^2 = N \sum k_i + 2 N \sum n_a k_i . \tag{5}$$

Substituting values from (5) and (2) in a standard formula for the standard deviation, one obtains

$$\sigma_E = \sqrt{\sum k_i + 2 \sum n_a k_i - (\sum k_i)^2}. \tag{6}$$

If all n items are of the same difficulty k, formulas (2) and (6) become, respectively,

$$\bar{E} = nk \tag{7}$$

and

$$\sigma_E = n\sqrt{k(1-k)}. \tag{8}$$

For one item, formulas (7) and (8) become

$$\bar{E} = k; \tag{9}$$

$$\sigma_E = \sqrt{k(1-k)}. \tag{10}$$

Correlations Between Sets of Items Whose Difficulties Are Known

It is now possible to find the expected Pearsonian correlation coefficient between two sets of items whose difficulties are known. Both sets of items are assumed to measure a single factor of ability, and the probability of chance success is zero.

Let $E_1 =$ an error score on item set 1, and $E_2 =$ an error score on item set 2. Then let

$$E_t \equiv E_1 + E_2 . \tag{11}$$

In this and in subsequent formulations, we shall let $N = 1$ for the purpose of simplification. This means that score frequencies and scatterplot cell frequencies are expressed as proportions of the population. (Even if N is left explicit, it drops out in the final formulas.) Then we may write, by (2),

$$\sum E_t = \sum^{n_1} k_1 + \sum^{n_2} k_2 , \tag{12}$$

and by (5),

$$\sum E_t^2 = \sum k_t + 2 \sum n_a k_t \,. \tag{13}$$

In determining values of n_a in (13), all items in both sets are arranged together in order of decreasing difficulty. In (12) and (13), the subscripts of k identify the k values as belonging to set 1, set 2, or the total (t) of the 2 sets. Squaring and summing equation (11), we obtain

$$\sum E_t^2 = \sum E_1^2 + \sum E_2^2 + 2 \sum E_1 E_2 \,,$$

or

$$\sum E_1 E_2 = \tfrac{1}{2}(\sum E_t^2 - \sum E_1^2 - \sum E_2^2). \tag{14}$$

Substituting expressions obtained in equations (5) and (13), we have

$$\begin{aligned} \sum E_1 E_2 &= \tfrac{1}{2}(\sum k_t + 2 \sum n_a k_t - \sum k_1 - 2 \sum n_a k_1 - \sum k_2 - 2 \sum n_a k_2) \\ &= \sum n_a k_t - \sum n_a k_1 - \sum n_a k_2 \,. \end{aligned} \tag{15}$$

The correlation r_{12} can be found by substituting the required values in the formula

$$r_{12} = \frac{\sum E_1 E_2 - (\sum E_1)(\sum E_2)}{\sqrt{[\sum E_1^2 - (\sum E_1)^2][\sum E_2^2 - (\sum E_2)^2]}} \,, \tag{16}$$

which gives

$$r_{12} = \frac{\sum n_a k_t - \sum n_a k_1 - \sum n_a k_2 - (\sum k_1)(\sum k_2)}{\sqrt{[\sum k_1 + 2 \sum n_a k_1 - (\sum k_1)^2][\sum k_2 + 2 \sum n_a k_2 - (\sum k_2)^2]}} \,. \tag{17}$$

Formula (17) shows that the correlation between two sets of items all of which measure a single ability can be written directly in terms of the number and difficulty of the items.

A number of interesting special cases of formula (17) can be written more simply. When all items of set 2 are harder than any item in set 1, the numerator of (17) can be written as $\sum k_1 (n_2 - \sum k_2)$. When each item in set 1 is matched by an item of corresponding difficulty in set 2, it will be found that the numerator of (17) becomes equal to the denominator. Hence, r_{12} in this case is equal to unity. This result is not surprising since the failure scores for the two sets of items should, in view of the assumptions outlined, exactly correspond.

When each set consists of only one item, we are actually dealing with the correlation of two items measuring a single ability. Let the subscript i refer to the easier item and j to the harder item. Still using failure scores and letting $N = 1$, we note that the cross-product term $\sum E_i E_j$ is equal to the proportion of individuals who fail

the easier item:

$$\sum E_i E_j = k_i . \tag{18}$$

The formula for the Pearsonian correlation coefficient for two items measuring a single factor can be developed, using formulas (9) and (10), as

$$\begin{aligned} r_{ij} &= \frac{\sum E_i E_j - \bar{E}_i \bar{E}_j}{\sigma_i \sigma_j} \\ &= \frac{k_i - k_i k_j}{\sqrt{k_i k_j (1 - k_i)(1 - k_j)}} \\ &= \sqrt{\frac{k_i (1 - k_j)}{k_j (1 - k_i)}} . \end{aligned} \tag{19}$$

Except for differences in notation, formula (19) is identical with a formula presented by Ferguson* as giving the value expected for the maximum correlation between two items homogeneous in content but not in difficulty. As Ferguson points out, formula (19) gives values of less than unity unless the items are equal in difficulty.

When each set consists of items of uniform difficulty, the error score for each set can be obtained by multiplying the error score for one of its component items by the number of items in the set, since all items in each set are passed or failed together. Such multiplication by a constant does not alter the relationship specified by formula (19) for any pair of items selected one from each set. Formula (19) therefore applies to this case, k_i and k_j referring to the difficulties of items in the easier and the harder sets, respectively.

Numerical Illustration of Formula (17)

In order to make concrete the operations involved in formula (17) and to relate them to conventional statistical procedures, a numerical example is given in Table 1. The analysis by formula (17) is shown in the left-hand portion of the table, while the conventional treatment is given at the right. Assume that we wish to find the expected correlation between two sets of items all of which measure a single ability. Set 1 consists of items with k values (proportions failing the item) of .45, .45, .45, .45, and .10, respectively; set 2 has items with k values of .95, .85, .85, .20, and .20, respectively. In Table 1 at (A), these items are arranged in order of decreasing difficulty, and

* Ferguson, G. A. The factorial interpretation of test difficulty. *Psychometrika*, 1941, 6, 323-329.

TABLE 1

Numerical Illustration of Formula (17) for Two Sets of Items

(A)

Computation from item difficulties

Set 1 k_1	Set 1 n_a	Set 2 k_2	Set 2 n_a
.45	0	.95	0
.45	1	.85	1
.45	2	.85	2
.45	3	.20	3
.10	4	.20	4

$\Sigma k_1 = 1.90$ $\qquad \Sigma k_2 = 3.05$

$\Sigma n_a k_1 = 3.10$ $\qquad \Sigma n_a k_2 = 3.95$

$\bar{E}_1 = 1.90$ $\qquad \bar{E}_2 = 3.05$

$\sigma_1^2 = 4.49$ $\qquad \sigma_2^2 = 1.6475$

(B)

Combination of set 1 and 2

k_t	n_a
.95	0
.85	1
.85	2
.45	3
.45	4
.45	5
.45	6
.20	7
.20	8
.10	9

$\Sigma k_t = 4.95$

$\Sigma n_a k_t = 14.55$

By formula (17),

$r_{12} = .6269\,.$

(C)

Computation from score distributions

Set 1 E_1	Set 1 f	Set 2 E_2	Set 2 f
0	.55	0	.05
1	.00	1	.10
2	.00	2	.00
3	.00	3	.65
4	.35	4	.00
5	.10	5	.20

$N = 1$ $\qquad N = 1$

$\Sigma E_1 = 1.90$ $\qquad \Sigma E_2 = 3.05$

$\Sigma E_1^2 = 8.10$ $\qquad \Sigma E_2^2 = 10.95$

$\bar{E}_1 = 1.90$ $\qquad \bar{E}_2 = 3.05$

$\sigma_1^2 = 4.49$ $\qquad \sigma_2^2 = 1.6475$

(D)

Combination of set 1 and 2

E_t	f
0	.05
1	.10
2	.00
3	.40
4	.00
5	.00
6	.00
7	.25
8	.00
9	.10
10	.10

$N = 1$

$\Sigma E_t = 4.95$

$\Sigma E_t^2 = 34.05$

By (14),

$$\Sigma E_1 E_2 = \tfrac{1}{2}(34.05 - 8.10 - 10.95) = 7.50.$$

By (16),

$$r_{12} = \frac{7.50 - (1.90)(3.05)}{\sqrt{[8.10 - (1.90)^2][10.95 - (3.05)^2]}} = .6269\,.$$

the values $\sum k_1$, $\sum k_2$, $\sum n_a k_1$, and $\sum n_a k_2$ are found. In order to find $\sum n_a k_t$, the items are completely rearranged in order of difficulty, as shown at (B), and the difficulty values are multiplied by the new values of n_a. The correlation coefficient is obtained by substitution in formula (17). At (C), the score distributions for the two sets are presented. The frequencies are expressed in proportions, and can be found as follows: the frequency of a failure score of zero is the proportion passing the hardest item; the frequency of a failure score of n is the proportion failing the easiest item; and the frequencies of the intermediate failure scores are the differences between the proportions failing adjacent items when the items are arranged in order of difficulty. At (D), the score distribution for the total failure score on set 1 and 2 is given, and the correlation coefficient is found by conventional methods after the term $\sum E_1E_2$ has been found by (14).

Table 2 shows the correlation surface for the two sets of items in

TABLE 2

Scatter-Diagram of Failure Scores on the Hypothetical Item Sets 1 and 2 Treated in Table 1. Cell values are proportions of the total population ($N = 1$)

		Failure Score, Set 1						
		0	1	2	3	4	5	Σ
Failure Score, Set 2	5	.00	.00	.00	.00	.10	.10	.20
	4	.00	.00	.00	.00	.00	.00	.00
	3	.40	.00	.00	.00	.25	.00	.65
	2	.00	.00	.00	.00	.00	.00	.00
	1	.10	.00	.00	.00	.00	.00	.10
	0	.05	.00	.00	.00	.00	.00	.05
	Σ	.55	.00	.00	.00	.35	.10	1.00

$r = .6269$

terms of proportions obtaining given score-combinations. The product-moment correlation coefficient obtained here is again .6269.

We have now analyzed the theoretical limiting case where all items measure a single factor. Relationships similar to those demonstrated here also hold in the case where we have two tests whose underlying continua are not perfectly correlated. The true correlation between two such tests can be expected to be obscured to some extent by differences in the difficulty levels of the two tests. Space does not

permit giving a complete demonstration of this fact, but it may be noted that given the side entries (frequencies) for any correlation table, the maximum positive product-moment correlation coefficient may be determined by considering that these entries fulfil the conditions of formula (17) and evaluating the formula. Thus, for example, given the side entries of Table 2, it is impossible to write a correlation table which will yield a correlation higher than .6269.

Statistics of a Set of Items When the Probability of Chance Success (c) Is Greater Than Zero

The formulations given thus far can be developed in such a way as to be applicable to the case where the probability of chance success is greater than zero. For example, the items may be cast in such a form that the individual may choose between several alternative responses. The probability of chance success (c) may if desired be determined on *a priori* grounds as the ratio of the number of correct alternatives to the total number of alternatives, but the method of determining c is irrelevant to the formulations given here. We shall set $d \equiv 1 - c$, for the sake of simplicity in certain expressions.

We shall first find the relations between (1) distributions of "true" failure scores unaffected by chance success and (2) distributions when the failure scores are affected by chance success. In the following, we do not need to assume that the items measure a single ability. However, it will be assumed that every subject who does not truly know an item will guess, *i. e.*, choose one of the alternatives.

Let us suppose we have a set of 5 items which are heterogeneous in difficulty and which are subject to passing by chance success. Let c be uniform for all items. Let $f_E \equiv$ the frequency of a true failure score unaffected by chance success. In Table 3, the actual failure scores of

TABLE 3

Frequency Distributions of Actual Failure Scores (E_c) Made by Those Obtaining Each True Failure Score (E)

	True Failure Score (E)						
E_c	0	1	2	3	4	5	Σ
0	f_0	cf_1	c^2f_2	c^3f_3	c^4f_4	c^5f_5	
1		df_1	$2cdf_2$	$3c^2df_3$	$4c^3df_4$	$5c^4df_5$	
2			d^2f_2	$3cd^2f_3$	$6c^2d^2f_4$	$10c^3d^2f_5$	
3				d^3f_3	$4cd^3f_4$	$10c^2d^3f_5$	
4					d^4f_4	$5cd^4f_5$	
5						d^5f_5	
Σ	f_0	f_1	f_2	f_3	f_4	f_5	N

those subjects with each true failure score are classified on the basis of the binomial theorem. For example, as shown in the column headed 0, those who make a true failure score of 0 do not guess and therefore make actual failure scores of 0. Those who make a true failure score of 1 have a chance c of passing the one remaining failed item to make an actual failure score of 0. The remainder, df_1, make an actual failure score of 1. The actual failure scores of those making a true failure score of 5 are distributed as shown in the column headed 5.

In order to find the mean ($\bar{E}_c$) and the standard deviation (σ_c) of the distribution of actual failure scores, we find the values of the sums $\sum E_c$ and $\sum E_c^2$. We can find these values for each column of Table 3 and sum over all the columns.

For any column of Table 3,

$$\sum E_c = f_E[0 \cdot c^E + 1 \cdot E\, c^{E-1} d + 2 \cdot \frac{E(E-1)}{1 \cdot 2} c^{E-2} d^2 + \cdots + E\, d^E]$$

$$= d\, f_E\, E[c^{E-1} + (E-1) c^{E-2} d + \frac{(E-1)(E-2)}{1 \cdot 2} c^{E-3} d^2 + \cdots + d^{E-1}]$$

$$= d\, f_E\, E\,. \tag{20}$$

Summing over all columns, in the general case,

$$\begin{aligned} \sum E_c &= d \sum f_E\, E \\ &= d \sum E\,. \end{aligned} \tag{21}$$

To find $\sum E_c^2$, we note that for any column of Table 3,

$$\sum E_c^2 = f_E[0^2 \cdot c^E + 1^2 \cdot E\, c^{E-1} d + 2^2 \cdot \frac{E(E-1)}{1 \cdot 2} c^{E-2} d^2 + \cdots + E^2 d^E]$$

$$= d\, f_E\, E[c^{E-1} + 2(E-1) c^{E-2} d + 3 \cdot \frac{(E-1)(E-2)}{1 \cdot 2} c^{E-3} d^2 + \cdots + E\, d^{E-1}],$$

or,

$$\sum E_c^2 = d\, f_E\, E\{ 0 \cdot c^{E-1} + [c^{E-1}] + (E-1) c^{E-2} d + [(E-1) c^{E-2}] + 2 \cdot \frac{(E-1)(E-2)}{1 \cdot 2} c^{E-3} d^2 + \left[\frac{(E-1)(E-2)}{1 \cdot 2} c^{E-3} d^2\right] + \cdots + (E-1) d^{E-1} + [d^{E-1}]\}. \tag{22}$$

In (22) the sum of the expressions within the braces but not in the square brackets can be found by a method similar to that employed in simplifying equation (20). The expressions within the square brackets represent the expansion of $(c + d)^{E-1} = 1$. Hence, for a given column of Table 3,

$$\begin{aligned}\sum E_c^2 &= df_E E[d(E-1)+1] \\ &= cdf_E E + d^2 f_E E^2 .\end{aligned} \tag{23}$$

Summing over all columns, we have

$$\begin{aligned}\sum E_c^2 &= c\, d \sum f_E E + d^2 \sum f_E E^2 \\ &= c\, d \sum E + d^2 \sum E^2 .\end{aligned} \tag{24}$$

Formulas for the mean and standard deviation of failure scores affected by chance can now be written in terms of the statistics of the distribution of "true" scores unaffected by chance, as follows:

$$\bar{E}_c = d\,\bar{E}; \tag{25}$$

$$\sigma_c = \sqrt{d^2\,\sigma_E^2 + c\,d\,\bar{E}} . \tag{26}$$

The Correlation between Actual Failure Scores on Two Sets of Items Affected by Chance Success When the Distributions of True or Non-Chance Scores Are Known

By (21) and (24) we can show the effect of chance success on the combined scores for set 1 and set 2. Let $E_{c_t} \equiv E_{c_1} + E_{c_2}$. By (21),

$$\sum E_{c_t} = d \sum E_t , \tag{27}$$

and by (24),

$$\sum E_{c_t}^2 = c\,d \sum E_t + d^2 \sum E_t^2 . \tag{28}$$

By analogy with (14),

$$\sum E_{c_1} E_{c_2} = \tfrac{1}{2}\left(\sum E_{c_t}^2 - \sum E_{c_1}^2 - \sum E_{c_2}^2\right) .$$

Substituting (24), expanding, and simplifying, we obtain,

$$\sum E_{c_1} E_{c_2} = d^2 \sum E_1 E_2 . \tag{29}$$

The expected correlation between E_{c_1} and E_{c_2} can be found by substituting the appropriate values in a formula for the Pearsonian product-moment correlation coefficient, as follows:

$$r_{c_1c_2} = \frac{\dfrac{\sum E_{c_1} E_{c_2}}{N} - \bar{E}_{c_1}\bar{E}_{c_2}}{\sigma_{c_1}\sigma_{c_2}}$$

$$= \frac{\dfrac{d^2 \sum E_1 E_2}{N} - d^2 \bar{E}_1 \bar{E}_2}{\sqrt{[d^2 \sigma_1^2 + c\,d\,\bar{E}_1][d^2 \sigma_2^2 + c\,d\,\bar{E}_2]}} \tag{30}$$

$$= \frac{d r_{12}\,\sigma_1\,\sigma_2}{\sqrt{d^2 \sigma_1^2 \sigma_2^2 + c\,d\,\bar{E}_1\,\sigma_2^2 + c\,d\,\bar{E}_2\,\sigma_1^2 + c^2 \bar{E}_1 \bar{E}_2}}.$$

Formula (30) holds whether or not the items measure a single ability.

Correcting Correlations for Chance Success

It may be of practical value to have formulas (25), (26), and (30) expressed in such a way that the "true" values $\bar{E}$, σ_E, and r_{12} can be estimated directly from the empirical values $\bar{E}_c$, σ_c, and $r_{c_1c_2}$. The required formulas are as follows:

$$\bar{E} = \frac{\bar{E}_c}{d}; \tag{31}$$

$$\sigma_E = \frac{\sqrt{\sigma_c^2 - c\,\bar{E}_c}}{d}; \tag{32}$$

$$r_{12} = \frac{r_{c_1c_2}\,\sigma_{c_1}\,\sigma_{c_2}}{\sqrt{\sigma_{c_1}^2 \sigma_{c_2}^2 - c\,\bar{E}_{c_1}\sigma_{c_2}^2 - c\,\bar{E}_{c_2}\sigma_{c_1}^2 + c^2 \bar{E}_{c_1}\bar{E}_{c_2}}}. \tag{33}$$

Thus, if we obtain the correlation between two multiple-choice tests of ability in which the subjects have responded to every item, we can estimate by (33) the correlation which would exist between the tests if the items were perfectly reliable and if the items were passed only by true mastery.

Where we have two equivalent forms of a test, *i. e.*, where $\bar{E}_{c_1} = \bar{E}_{c_2}$ and $\sigma_{c_1} = \sigma_{c_2}$, formula (33) becomes

$$r_{12} = \frac{r_{c_1c_2}\,\sigma_{c_1}^2}{\sigma_{c_1}^2 - c\,\bar{E}_{c_1}}. \tag{34}$$

Effect of Chance Success on Sets of Items Measuring a Single Ability

We now return to the analysis of items measuring a single ability. Formulas (25), (26), and (30) can be written directly in terms of the true difficulties (k values) of the items concerned by substituting formulas (2), (6), and (17), respectively:

$$\bar{E}_c = d\,\bar{E} = d\sum^{n} k_i\,; \tag{35}$$

$$\begin{aligned} \sigma_c{}^2 &= d^2\,\sigma_E{}^2 + c\,d\,\bar{E} \\ &= d\sum k_i + 2\,d^2 \sum n_a\,k_i - d^2(\sum k_i)^2\,; \end{aligned} \tag{36}$$

$$\begin{aligned} r_{c_1c_2} &= \frac{d\,r_{12}\,\sigma_1\,\sigma_2}{\sqrt{d^2\,\sigma_1{}^2\,\sigma_2{}^2 + c\,d\,\bar{E}_1\,\sigma_2{}^2 + c\,d\,\bar{E}_2\,\sigma_1{}^2 + c^2\,\bar{E}_1\,\bar{E}_2}} \\ &= \frac{d[\sum n_a\,k_t - \sum n_a\,k_1 - \sum n_a\,k_2 - (\sum k_1)(\sum k_2)]}{\sqrt{[\sum k_1 + 2\,d\sum n_a\,k_1 - d(\sum k_1)^2][\sum k_2 + 2\,d\sum n_a\,k_2 - d(\sum k_2)^2]}}\,. \end{aligned} \tag{37}$$

Special cases of formula (37) can now be written by introducing modifications of formula (17) as before. When all items of set 2 are harder than any item in set 1,

$$r_{c_1c_2} = \frac{d\sum k_1(n_2 - \sum k_2)}{\sqrt{[\sum k_1 + 2\,d\sum n_a\,k_1 - d(\sum k_1)^2][\sum k_2 + 2\,d\sum n_a\,k_2 - d(\sum k_2)^2]}}\,. \tag{38}$$

When each item in set 1 is matched by an item of corresponding difficulty in set 2, $r_{12} = 1.00$, $\bar{E}_1 = \bar{E}_2$, and $\sigma_1 = \sigma_2$. Hence,

$$r_{c_1 c_2} = \frac{d[\sum k_1 + 2\sum n_a\,k_1 - (\sum k_1)^2]}{\sum k_1 + 2\,d\sum n_a\,k_1 - d(\sum k_1)^2}\,. \tag{39}$$

When each set consists of items uniform in difficulty, we have for any one of the sets, by (25) and (7),

$$\bar{E}_c = d\,\bar{E} = d\,n\,k\,; \tag{40}$$

and by (26) and (8),

$$\begin{aligned} \sigma_c{}^2 &= d^2\,\sigma_E{}^2 + c\,d\,\bar{E} \\ &= d^2\,n^2\,k(1-k) + c\,d\,n\,k\,. \end{aligned} \tag{41}$$

By (29), and in view of the fact that the items are uniform in difficulty,

$$\sum E_{c_1} E_{c_2} = d^2 \sum E_1 E_2 = d^2 n_2 \sum k_1 = d^2 n_2 n_1 k_1, \tag{42}$$

where $N = 1$ and the subscript 2 refers to the harder set of items. Hence,

$$\begin{aligned} r_{c_1 c_2} &= \frac{d^2 n_2 n_1 k_1 - (d n_1 k_1)(d n_2 k_2)}{\sqrt{[d^2 n_1^2 k_1(1-k_1) + c\, d\, n_1 k_1][d^2 n_2^2 k_2(1-k_2) + c\, d\, n_2 k_2]}} \\ &= \frac{d\, n_1 n_2 k_1 (1-k_2)}{\sqrt{[n_1 k_1(c + d n_1 - d n_1 k_1][n_2 k_2(c + d n_2 - d n_2 k_2)]}}. \end{aligned} \tag{43}$$

Where $n_1 = n_2$ formula (43) becomes

$$r_{c_1 c_2} = \frac{d\, n\, k_1 (1-k_2)}{\sqrt{k_1 k_2 (c + d n - d n k_1)(c + d n - d n k_2)}}. \tag{44}$$

Further, when $k_1 = k_2$, that is, where the tests are of exactly the same difficulty and may be regarded as "equivalent" forms,

$$r_{c_1 c_2} = \frac{d\, n(1-k)}{c + d\, n(1-k)}. \tag{45}$$

It can be shown that the correlation estimated by (45) varies in accordance with the Spearman-Brown formula for lengthened test-reliability. Let ν be the number of times a test of n items is increased in length; let $r_{\nu\nu}$ be the reliability of a test of length νn. By the Spearman-Brown formula,

$$r_{\nu\nu} = \frac{\nu\, r_{nn}}{1 + (\nu - 1) r_{nn}}.$$

Substituting formula (45) for r_{nn}, we obtain

$$\begin{aligned} r_{\nu\nu} &= \frac{\dfrac{\nu\, d\, n(1-k)}{c + d\, n(1-k)}}{1 + (\nu - 1)\dfrac{d\, n(1-k)}{c + d\, n(1-k)}} \\ &= \frac{\nu\, d\, n(1-k)}{c + \nu\, d\, n(1-k)}. \end{aligned} \tag{46}$$

If we consider n as a variable quantity, νn is merely a particular value of n; hence (46) is essentially equivalent to formula (45), and it is shown that the reliability of a test as estimated by formula (45) varies as a function of the length of the test according to the Spearman-Brown prophecy formula.

When each set consists of one item, formula (44) becomes

$$r_{12} = \frac{d\,k_1(1-k_2)}{\sqrt{k_1\,k_2(1-d\,k_1)\,(1-d\,k_2)}}. \tag{47}$$

Application of the Tetrachoric Correlation Coefficient

Thus far all formulations presented here have been in terms of the Pearsonian product-moment correlation coefficient. There is considerable question as to whether this statistic is applicable in all cases which have been treated. Its use in evaluating the relation between two items appears to be inappropriate in view of the broad categories involved. The Pearsonian coefficient affords a means of estimating the efficiency of prediction of the score on one item from the score on another item. If the primary concern is not with prediction, however, but with the factorial relation between two items, the Pearsonian coefficient does not give a correct indication of this relation because even where sets of items measure a single ability, the correlation coefficient varies widely as a function of the difficulties of the items.

FIGURE 1

Abac Showing the Expected Tetrachoric Correlations Between Items Measuring a Single Ability, When $c_i = c_j = .50$. Values above the diagonal line are meaningless.

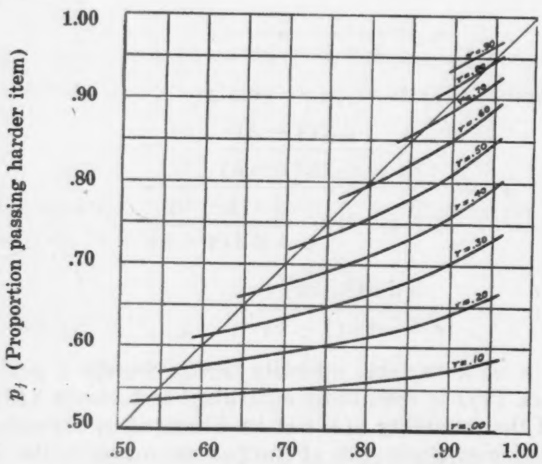

TABLE 4

Correlation Surface Expected for Two Sets of Items All Measuring a Single Ability, When $k_1 = .4$, $k_2 = .8$, and $c = .5$ throughout, the Items in Each Test Being Uniform in Difficulty

		Failure Score, Set 1						
		5	4	3	2	1	0	Σ
Failure Score, Set 2	0	.000390625	.001953125	.003906250	.003906250	.001953125	.212890625	.225000000
	1	.001953125	.009765625	.019531250	.019531250	.009765625	.064453125	.125000000
	2	.003906250	.019531250	.039062500	.039062500	.019531250	.128906250	.250000000
	3	.003906250	.019531250	.039062500	.039062500	.019531250	.128906250	.250000000
	4	.001953125	.009765625	.019531250	.019531250	.009765625	.064453125	.125000000
	5	.000390625	.001953125	.003906250	.003906250	.001953125	.012890625	.025000000
	Σ	.012500000	.062500000	.125000000	.125000000	.062500000	.612500000	1.000000000

Numerical evaluation of the various formulas presented in this paper will demonstrate that (a) the obtained correlation coefficient of tests or items decreases as the tests or items become less similar in difficulty; and (b) other things being equal, the obtained correlation of pairs of items decreases as their average difficulty becomes greater.

The tetrachoric correlation coefficient may be applicable to some of the cases which have been treated. Given the frequencies in a 2×2 table, we can estimate by means of r_{tetr} the degree of correlation represented by the best normal correlation surface fitted to these frequencies. r_{tetr} is thus an estimate of the correlation between the continua underlying two items or two sets of items. When chance success is not operating, and when the items measure the same ability, r_{tetr} will always be unity, because however the variables are dichotomized, one cell in the 2×2 table will be vacant. r_{tetr} may profitably be used as an indication of the extent to which a pair of items fall on a homogeneous continuum of difficulty and overlap factorially.

Where chance success is a factor (*i. e.*, where $c > .00$), the tetrachoric correlation tends to vary in the same manner as the Pearsonian correlation; *i. e.*, the obtained correlation coefficient tends to decrease as the tests become less similar in difficulty. Figure 1 is an abac showing for illustrative purposes the inter-item tetrachoric correlations expected when the items are assumed to measure a single ability and when $c = .5$ for both items.

When the measurements correlated consist of more than one item and are uniform in difficulty, the value of the tetrachoric correlation is markedly affected by the point of dichotomization in each variable. For purposes of illustration, Table 4 shows the correlation surface that would result if we had two sets of items all measuring a single factor, where each set is of uniform difficulty and where each item is

TABLE 5

Tetrachoric Correlations Obtained by Various Dichotomizations of Distributions in Table 4

Set 2—Dichotomization between:	Set 1—Dichotomization between: 5—4	4—3	3—2	2—1	1—0
1—0	*	.50	.60	.68	.75
2—1	*	.25	.33	.40	.45
3—2	*	.13	.18	.23	.27
4—3	*	.07	.11	.15	.18
5—4	*	*	*	*	*

* Values cannot be determined from Thurstone tables on account of small side-entry values.

affected by chance success. Test 1 is assumed to be composed of items with difficulties (k) of .4; for test 2, $k = .8$. These difficulties are in terms of the proportions who fail the items when there is no possibility of chance success. It is assumed that $c = .5$ throughout. The frequencies, expressed as proportions, have been computed by (a) writing the cell-frequencies in the non-chance situation, and (b) distributing by *a priori* probability the chance-affected failure scores of those who are item failers in the non-chance situation. The Pearsonian correlation for the table is .25 by formula (44); this value can also be found directly from the table by conventional procedures. Table 5 shows the tetrachoric correlations which are obtained by taking all the possible pairs of dichotomization points. Dichotomization as near the medians as possible (between scores of 0 and 1 for test 1 and between scores of 2 and 3 for test 2) yields $r_{tetr} = .27$. The highest tetrachoric obtainable, .75, is for dichotomization between scores of 0 and 1 for both tests.

If it is desired to use the tetrachoric correlation to estimate the factorial relation between items where chance success can operate, it is necessary to correct the proportions in the 2×2 table for the effect of chance success. In order to minimize random error, the correction should not be attempted unless a large number of cases (say, 200 or more) are available and unless the value of c can be estimated with considerable confidence. This correction is made as follows:

(1) Correct the side entries for chance success. Let $p_i =$ the proportion of individuals actually passing an item, whether by true mastery or by chance. k_i is the estimated proportion who would fail the item when the factor of chance success is not operating. Then

$$p_i = (1 - k_i) + c_i\, k_i = 1 - d_i\, k_i; \tag{48}$$

$$k_i = \frac{1 - p_i}{d_i} = \frac{q_i}{d_i}. \tag{49}$$

(2) Estimate the corrected value of the proportions inside the 2×2 table. Let $q_{ij} =$ the proportion actually failing both items; let $k_{ij} =$ the estimated proportion failing both items when guessing is not possible. Then by *a priori* probability,

$$q_{ij} = d_i\, d_j\, k_{ij}; \tag{50}$$

$$k_{ij} = \frac{q_{ij}}{d_i\, d_j}. \tag{51}$$

(3) Fill out the remainder of the cells by subtraction.

For example, suppose a four-fold table (Table 6) has been obtained for two items where $c_1 = c_2 = .25$. The tetrachoric correlation determined from Table 6 is .29. In order to correct this table for chance, we construct Table 7. By (49), the estimated k values of items 1 and 2 are .467 and .600, respectively. By (51), the estimated true proportion of persons failing both items is $.20/(.75)^2 = .356$. The tetrachoric correlation estimated from Table 7 is now .48.

TABLE 6
Proportions of Individuals Passing and Failing Two Items
Where Chance Success is Possible

		Item 1 Fail	Item 1 Pass	Total
Item 2	Pass	.15	.40	.55
	Fail	$.20 = q_{12}$	.25	$.45 = q_2$
	Total	$.35 = q_1$	.65	1.00

$r_{tetr} = .29$

TABLE 7
Proportions of Individuals Passing and Failing Two Items
by Actual Mastery Alone, Estimated from Table 6

		Item 1 Fail	Item 1 Pass	Total
Item 2	Pass	.111	.289	.400
	Fail	.356	.244	.600
	Total	.467	.533	1.000

$r_{tetr} = .48$

It will sometimes happen that k_{ij} as estimated by (51) is greater than one or both of the values k_i or k_j. This result may represent a random deviation from theoretical probability or it may be due to an overestimation of c_i or c_j. In practice, it will probably be found best to infer from this result a correlation approaching unity unless there is external evidence that either c_i or c_j has been grossly overestimated.

In passing it may be noted that formulas (49) and (51) are special cases of (21) and (29), respectively.

Implications

It has been shown that Pearsonian correlations between items or sets of items measuring abilities tend to vary as a function of (a) the difficulties of the items involved, and (b) the extent to which chance success by guessing is possible. Several means of estimating the true factorial overlap between items are proposed, *viz.*, the use of the tetrachoric correlation coefficient and the correction of 2×2 tables for chance by *a priori* probability theory. These techniques can be applied only under rather severely circumscribed conditions; for example, it is necessary that all subjects shall have made some response to each item involved. Nevertheless, it may be found profitable to set up rigorously controlled testing conditions in order to take advantage of the formulations presented in this paper.

Correlations between items or small sets of items uniform in difficulty which have appeared in the literature must be carefully scrutinized for the possibility that they may have been subject to spurious influences such as those with which we have dealt here. Factorial studies of items must be examined for the possibility that heterogeneity of items in difficulty has given rise to spurious factors.

Techniques for studying the factorial composition of items are not yet adequate. It will quite probably be found that the factor analysis of items cannot be based upon any correlation measure which can be derived from a pair of items alone, since all that can be established from a pair of items alone is the extent to which they measure on a "polar" scale of difficulty such that all who pass the harder item also pass the easier and all who fail the easier item also fail the harder.

PSYCHOMETRIKA—VOL. 10, NO. 1
MARCH, 1945

INTERPRETATION OF SECOND-ORDER FACTORS

KARL J. HOLZINGER
UNIVERSITY OF CHICAGO

It is shown that a "second-order" factor pattern is equivalent to the transformation employed in rotating an orthogonal factor pattern to an oblique form. The correlation among the second-order factors may then be interpreted as due to the original first-order factors.

The question is often raised, "What has become of the general orthogonal factor when an oblique solution is made?" The answer generally is that the general factor is expressed somehow by the intercorrelations of the oblique factors. The present paper is concerned with giving more precise answers to such questions.

When correlations are employed in factor analysis, an oblique solution can be made only by a transformation of an orthogonal solution already known. Since we are concerned here only with common factors, this transformation will be made in the common-factor space (abbreviated hereafter as c.f.s.). The analysis which follows will be illustrated at each stage with the seven-variable bi-factor pattern from Holzinger's *Manual.* This example is chosen to illustrate the relationship between the general factor and the intercorrelations of the oblique factors, but any other pattern would have served equally well for the general problem.

Let the common-factor portion of an orthogonal factor pattern be written in the form,

$$Z_j = A_{js} F_s \,, \tag{1}$$

where the common factors $F_s (s = 1, 2, 3, \cdots, m)$ are in standard form and A_{js} is the matrix (of rank m) of their coefficients for the variables $Z_j (j = 1, 2, 3, \cdots, n)$, which are not in standard form. The variables Z_j are the projections in the c.f.s. of the entire variables which also include the unique factors. The illustration of equation (1) may be written as follows:

$$\begin{aligned}
Z_1 &= .7F_1 + .5F_2 + 0\\
Z_2 &= .6F_1 + .6F_2 + 0\\
Z_3 &= .5F_1 + .8F_2 + 0\\[1ex]
Z_4 &= .7F_1 + 0 + .6F_3 \quad \text{rank} = 3\\
Z_5 &= .6F_1 + 0 + .6F_3\\
Z_6 &= .4F_1 + 0 + .8F_3\\[1ex]
Z_7 &= .8F_1 + 0 \quad + 0\,.
\end{aligned} \tag{1}'$$

An oblique solution may be obtained from equation (1) by means of a transformation derived from the composite variables,

$$\begin{aligned}
V_1 &= C_{11}F_1 + C_{12}F_2 + \cdots + C_{1m}F_m\\
V_2 &= C_{21}F_1 + C_{22}F_2 + \cdots + C_{2m}F_m\\
&\ldots\ldots\ldots\ldots\ldots\\
V_m &= C_{m1}F_1 + C_{m2}F_2 + \cdots + C_{mm}F_m\,.
\end{aligned} \tag{2}$$

These composite variables are formed by adding the equations within the m subgroups member by member. For the illustrative example, therefore,

$$\begin{aligned}
V_1 &= 1.8F_1 + 1.9F_2 + 0\\
V_2 &= 1.7F_1 + 0 + 2.0F_3\\
V_3 &= .8F_1 + 0 + 0\,.
\end{aligned} \tag{2}'$$

The standard deviations of the composite variables (in the c.f.s) is given by the formula

$$\sigma_s = \sqrt{C^2_{s1} + C^2_{s2} + \cdots + C^2_{sm}}\,, \tag{3}$$

which involves only the C's for uncorrelated factors F_s. For the illustrative example,

$$\begin{aligned}
\sigma_1^2 &= (1.8)^2 + (1.9)^2 = 6.85; &\quad \sigma_1 &= 2.617; &\quad \frac{1}{\sigma_1} &= .382\\
\sigma_2^2 &= (1.7)^2 + (2.0)^2 = 6.89; & \sigma_2 &= 2.625; & \frac{1}{\sigma_2} &= .381\\
\sigma_3^2 &= (.8)^2; & \sigma_3 &= .800; & \frac{1}{\sigma_3} &= 1.250\,.
\end{aligned} \tag{3}'$$

The composite variables (2) are next changed to standard form by dividing V_s by σ_s to give the equations,

$$L_s = \frac{V_s}{\sigma_s}. \tag{4}$$

For the illustration these equations become

$$
\begin{aligned}
L_1 &= \frac{1.8}{2.617}F_1 + \frac{1.9}{2.617}F_2 + 0 &&= .688F_1 + .726F_2 + 0 \\
L_2 &= \frac{1.7}{2.625}F_1 + 0 + \frac{2.0}{2.625}F_3 &&= .648F_1 + 0 + .762F_3 \\
L_3 &= \frac{.8}{.8}F_1 + 0 + 0 &&= F_1 + 0 + 0.
\end{aligned}
\tag{4)'}
$$

The standardized variables L_s are defined here as the oblique factors; that is, they are the composite variables V_s standardized in the c.f.s.

The coefficients in equation (4) form a transformation matrix employed when an oblique solution is desired.* The rows of this matrix are the direction cosines of the oblique factors L_s with respect to the original orthogonal factors F_s.

Equation (4) is also a second-order factor pattern, the coefficients of the factors being given by the columns. The correlation amongst the factors L_s may therefore be interpreted as due to the original orthogonal factors F_s as illustrated by equations (4)′. The general factor in equations (1)′ is the same as the general factor in equations (4)′, but the group factors in (1)′ are represented in equations (4)′ as unique factors in the c.f.s. This type of uniqueness will occur only when A_{js} is of bi-factor form.

In the above analysis it was assumed that the number of composite variables is equal to the rank of the matrix (1). This assumption is necessary in order to obtain exact relationships between factors in the same space. In case only $m-1$ composites are employed, equations of the form (4) may be obtained, but the relationships shown will be only approximate because different spaces are involved.

If a bi-factor pattern with group factors for all variables is employed, the number of composite variables will usually be one less than the number of common factors. It has sometimes been argued that the rank determined by the composites is the correct one and that the extra factor in the bi-factor pattern is therefore unwarranted. This reasoning is very circular, because we never know the exact rank with actual data involving several factors. It might be argued

* The oblique solution consists of the structure S_{js}, which gives the correlations between tests and factors, and the pattern B_{js}, which represents the coefficients in the linear expressions between tests and oblique factors. The structure is obtained from the above transformation T_{ss} and the pattern A_{js} by the equation $A_{js}T_{ss} = S_{js}$. The pattern B_{js} is then obtained from the equation $B_{js} = S_{js}\phi_{ss}^{-1}$, where θ_{ss} is the matrix of the intercorrelations of the factors L_s.

equally well that the rank of the oblique solution is lower than the data warrant.

A more general treatment of the above problem will next be given. After the composite variables have been chosen and the σ_s computed, the above transformation matrix, coefficients of equation (4), may be expressed in the form $P_{sj}A_{js}$, where

$$P_{sj} = \begin{Vmatrix} \frac{1}{\sigma_1} & \frac{1}{\sigma_1} & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\sigma_2} & \frac{1}{\sigma_2} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \frac{1}{\sigma_m} \end{Vmatrix} . \tag{5}$$

Equation (4) may then be written in the form,

$$L_s = [P_{sj}A_{js}]F_s , \quad \text{or} \quad L = (PA)F. \tag{6}$$

For the illustrative example the product $P_{sj}A_{js}$ is

$$\begin{Vmatrix} .382 & .382 & .382 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .381 & .381 & .381 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.25 \end{Vmatrix} \times \begin{Vmatrix} .7 & .5 & 0 \\ .6 & .6 & 0 \\ .5 & .8 & 0 \\ .7 & 0 & .6 \\ .6 & 0 & .6 \\ .4 & 0 & .8 \\ .8 & 0 & 0 \end{Vmatrix} = \begin{Vmatrix} .688 & .726 & 0 \\ .648 & 0 & .762 \\ 1 & 0 & 0 \end{Vmatrix} ,$$

the matrix on the right being the coefficients in equations (4)′. The transformation PA thus sections the matrix (1) into m composite variables and standardizes these composites in one step, giving the required form of the transformation matrix for future analysis.

It is of some interest to relate the above transformation $P_{sj}A_{js}$ to one more generally employed. Let the oblique pattern for all j variables be written in the form,

$$Z_j = B_{js}L_s , \tag{7}$$

or more simply ,

$$Z = BL . \tag{8}$$

Assuming that patterns (1) and (8) are known, the problem is to show the relationships between factors. For an oblique pattern B and an orthogonal pattern A, the problem is to find a matrix T such that

$$BT = A\,. \tag{9}$$

Since B does not ordinarily have an inverse we cannot simply write

$$T = B^{-1}A\,. \tag{10}$$

Noting, however, that

$$(B'B)^{-1}(B'B) = I\,, \tag{11}$$

and premultiplying both sides of (9) by $(B'B)^{-1}B'$, we find

$$T = (B'B)^{-1}B'A\,, \tag{12}$$

which is the form generally employed.

It will next be shown that PA of equation (6) is identical with T. Substituting the right member of (6) in equation (8) gives

$$Z = BPAF\,. \tag{13}$$

From equations (1) and (13) it is apparent that

$$A = BPA\,. \tag{14}$$

Substituting the right member of (14) in (12) gives

$$T = (B'B)^{-1}(B'B)PA = IPA = PA\,. \tag{15}$$

The transformation PA is thus equivalent to T but is obviously much simpler.

The answer to the question at the beginning of this paper would appear to be that the correlation amongst factors may be interpreted as an expression of the original first-order factors. The factorial job is done, however, when a solution of the form (1) or (8) and its accompanying structure are obtained. It is therefore not recommended that second-order factors be explicitly found in the hope of discovering something psychologically new. Such factors merely furnish a statistical interpretation of the aspects of (1) and (8).

PSYCHOMETRIKA—VOL. 10, NO. 1
MARCH, 1945

A NOTE ON STEADY STATES AND THE WEBER-FECHNER LAW

S. SPIEGELMAN
WASHINGTON UNIVERSITY

J. M. REINER
UNIVERSITY OF MINNESOTA

The steady state of a simple reaction system has been shown to have some of the properties of a psychophysical discrimination system, including the possibility of deducing a generalized Weber-Fechner Law, both in integral form and in difference form. The Weber ratio so deduced is not constant, and its dependence on stimulus intensity is exhibited. The dependence of the difference limen on the internal threshold is discussed; it is found that in general there is a finite value of this threshold for which response is impossible. This critical threshold is lower for higher values of the reference stimulus intensity. Similarly, it is shown that the difference limen and the Weber ratio, for a fixed value of the threshold, become infinite (i.e., discrimination is impossible) for a value of the stimulus intensity which in general is finite.

In the course of investigating the theoretical properties of the steady state in biological systems, the authors found a relation which is strongly suggestive of the Weber-Fechner law. This analogy has been mentioned by Burton (1), unfortunately without proof or elaboration; Burton states that "The mathematical proof is complicated." The mathematical development is in fact rather simple; and it is the purpose of this note to present it, and to indicate some of the consequences which may be deduced from this relation.

The theoretical model in question is a physical system in which a substance A appears from a source S. A is transformed chemically into another substance B, which disappears into a "sink" Z. (The source and sink may be conceived of as the environment of the physical system, from and into which the substances A and B may diffuse.)

Let the concentrations of A and B and of the source and sink be represented by the letter c with appropriate subscripts. Then the variation of A and B with time is prescribed by the following pair of differential equations:

$$\frac{dc_A}{dt} = k_0 c_S - (k_0 + k) c_A + k' c_B;$$
$$\frac{dc_B}{dt} = k c_A - (k' + k_Z) c_B + k_Z c_Z. \tag{1}$$

In the above, k_0 and k_Z measure the rate of diffusion between the system and the source or sink, respectively, while k and k' are the velocity constants (2, p. 952) for the transformation of A into B and B into A, respectively.

These equations may be solved for c_A and c_B as functions of the time t. It is then easily shown that these functions, with increasing time, always approach stationary values. However, these values may be found without solving the differential equations. For, when the concentrations are stationary and no longer change with time, it follows that:

$$\frac{dc_A}{dt} = 0; \quad \frac{dc_B}{dt} = 0. \tag{2}$$

Accordingly, we may fulfill this condition by setting the right-hand sides of the differential equations equal to zero; this gives a pair of simultaneous linear equations which may be solved for c_A and c_B. The solutions are:

$$c_A = \frac{k_0 c_S (k' + k_Z) + k' k_Z c_Z}{k_0 (k' + k_Z) + k k_Z};$$
$$c_B = \frac{k_Z c_Z (k_0 + k) + k_0 c_S k}{k_0 (k' + k_Z) + k k_Z}. \tag{3}$$

Let us now consider a possible interpretation of this simple model. Suppose that an external stimulus acting upon a biological system increases the rate at which A is transformed into B; we might then take k as a measure of the stimulus intensity. Suppose further that the level of the organism's response is determined by the concentration of the product B; we might then take c_B as a measure of the response. If we suppose that measurements are made under stationary conditions (i.e., that the time required to attain the stationary state is short compared with the time of observation), then the second of equations (3), giving c_B as a function of k, represents the relation between stimulus and response.

According to Burton (1, p. 333), the Weber-Fechner law states that "progressive equal increments of intensity of stimulus produce decreasing increments of response." In terms of our symbols, this

would mean that the plot of c_B against k has progressively decreasing slope. In order to prove that equation (3) satisfies this requirement, it is only necessary to show that the second derivative of c_B with respect to k is everywhere negative. This follows immediately by differentiating twice:

$$\frac{d^2c_B}{dk^2} = -2\,k_Z\,\frac{k_0[k_0\,c_S(k'+k_Z)+k'\,k_Z\,c_Z]}{[k_Z\,k+k_0(k'+k_Z)]^3}. \tag{4}$$

Since all the quantities in the above expression are intrinsically positive, it is clear that this expression is everywhere negative.

According to the classical formulation of Fechner, c_B should be a linear function of $\log k$. However, it has long been known that data obtained by psychophysical measurements fulfill such a relation only over a limited middle range of stimulus intensity; when the intensity scale is extended in both directions, the data usually fit an ogive or S-shaped curve, which has an approximately linear portion in the neighborhood of the inflection point near the middle (3, 4, p. 114). It is therefore of some interest to show that equation (3) also meets with this condition. For the sake of simplicity in writing the formulas, we introduce the following abbreviations:

$$\begin{aligned} a &= k_0\,c_S + k_Z\,c_Z \\ b &= k_0\,k_Z\,c_Z \\ c &= k_Z \\ d &= k_0(k'+k_Z) \\ u &= \log k\,. \end{aligned} \tag{5}$$

Equation (3) then takes the form

$$c_B = \frac{a\,e^u + b}{c\,e^u + d}. \tag{6}$$

The first and second derivatives with respect to u are:

$$\frac{dc_B}{du} = \frac{(a\,d - b\,c)\,e^u}{(c\,e^u + d)^2}; \tag{7}$$

$$\frac{d^2c_B}{du^2} = \frac{(a\,d - b\,c)\,(d - c\,e^u)\,e^u}{(c\,e^u + d)^3}. \tag{8}$$

From these equations we can easily deduce the properties of the curve. From equation (6) we see that, as u approaches negatively infinite values, c_B approaches the constant value b/d; as u approaches positively infinite values, c_B approaches the constant limit a/c. More-

over, from the definitions (5) it is easily shown that b/d is less than a/c.

Examining equation (7), we note that $ad - bc$ is a positive quantity according to (5). The first derivative is therefore always greater than or equal to zero. Letting u approach positive and negative infinity, we find that the slope is equal to zero at both limits.

From equation (8) we may find the inflection point by setting the expression for the second derivative equal to zero. There is only one finite solution, whose value is log (d/c).

Thus we have a curve in which the response passes from a positive plateau upwards through an inflection point to another plateau. This is precisely the description of an ogive.

So far we have, among other things, given a proof for the assertions offered without proof by Burton. However, a relation between stimulus intensity and the corresponding level of response is not precisely what was contemplated in the original formulation of the Weber-Fechner law. Many of the relevant experiments are concerned rather with the relation between *increments* of stimulus and response, and in particular with finite increments rather than differentials. It seems of value, therefore, to see what our theoretical model yields in such a situation.

Let the (in general finite) increment of k be denoted by δk; let the corresponding increment of c_B be denoted by δc_B. We obtain the expression for δc_B by using equation (3), with $k + \delta k$ substituted for k, and subtracting from this the original form of equation (3). We have then

$$\delta c_B = k_0 \, \delta k \frac{k_0 \, c_S \, k(k' + k_Z) + k' \, k_Z \, c_Z}{[k_0(k' + k_Z) + k_Z \, k][k_0(k' + k_Z) + k_Z \, k + k_Z \, \delta k]} . \tag{9}$$

This may be compared with the original formulation of Fechner, which in our notation would give (4)

$$\delta c_B = K \frac{\delta k}{k} . \tag{10}$$

In our case, not only is the response increment not simply inversely proportional to k, but it depends on the stimulus increment in a more complicated way than that given by direct proportionality. In particular, our formulation predicts that, for sufficiently large increments of stimulus intensity, the increment of response approaches a constant limiting value which is independent of the stimulus increment. In other words, if the jump from one stimulus intensity to a

new one be sufficiently large, the responding organism can tell that the new stimulus is larger, but no longer can estimate how much larger it is.

It may be objected to this, as well as to most of our formulation, that it implies the ability of subjects in psychophysical discrimination experiments to make quantitative estimates of their perception of stimulus intensities or stimulus differences, and that in fact they can only make such judgments as "greater," "less," or "equal." We would point out to such critics that this is chiefly a consequence of the way in which such experiments are set up and in which the types of judgments to be given are determined by the observer. Naturally, the accuracy of such judgments by subjects will vary considerably with their acquaintance with the kind of magnitude concerned, and their consequent possession of what one might call an internal scale of measurement; but this does not at all affect the principle of our quantitative model. To anyone who doubts that human beings can ever estimate magnitudes quantitatively, we recommend observation of a surveyor or carpenter estimating lengths, or a butcher measuring out a pound of hamburger.

Let us consider now the situation in which the stimulus increment may be characterized as a just noticeable difference. It is implicit in the theoretical description of such a situation that the substance B which determines the response does not produce a discriminating response for every increment by which it may increase, but only if the increment equals or exceeds a certain threshold value. We want now to find the stimulus increment which corresponds to this threshold; this would be the just noticeable difference of stimulus.

We call the threshold T. Let us substitute T for $\delta\, c_B$ in equation (9), and solve for $\delta\, k$. This gives

$$\delta\, k = \frac{T[k_0(k' + k_Z) + k_Z\, k]^2}{k_0[k_0\, c_S(k' + k_Z) + k'\, k_Z\, c_Z] - T\, k_Z[k_0(k' + k_Z) + k_Z\, k]}. \tag{11}$$

The difference limen δk increases with k as should be the case (4, p. 126, p. 163).

The most obviously striking thing about this relation is the way in which it exhibits the effect of imposing a threshold on the substance B. For as T increases, δk does not merely increase more or less rapidly with it, as one might at first expect. There is in addition a finite value of T for which δk becomes infinite. That is to say, if the discrimination threshold of an organism reaches a certain point, the organism can no longer discriminate any finite change of stimu-

lus, no matter how large. This limiting value of the threshold is given by that value of T for which the denominator of equation (11) vanishes, namely,

$$T = k_0[k_0\, c_S(k' + k_Z) + k'\, k_Z\, c_Z] \Big/ k_Z[k_0(k' + k_Z) + k_Z\, k]\,. \qquad (12)$$

This value is a function of the previously existing stimulus level k. As may be seen from equation (12), it varies inversely with k; the higher the previous stimulus intensity, the lower is the threshold value required to block completely the discrimination of a new stimulus intensity. If the initial stimulus is sufficiently large, the discrimination is impossible for an organism with any finite threshold, no matter how small.

The value of δk for $k = 0$ is the stimulus limen or RL (4, p. 111).

It is instructive to write equation (12) in a form which explicitly gives $\delta k / k$, the form perhaps most familiar in the psychophysical literature as the Weber ratio. This function in our case is

$$\frac{\delta k}{k} = \frac{T[k_0(k' + k_Z) + k_Z\, k]^2}{k\{k_0[k_0\, c_S(k' + k_Z) + k'\, k_Z\, c_Z] - T\, k_Z[k_0(k' + k_Z) + k_Z\, k]\}}. \qquad (13)$$

For small values of k, the function is very large, and it decreases as k increases. However, the denominator vanishes and the function becomes infinite at a finite value of k given by

$$k = \frac{k_0}{k_Z}\, \frac{k_0\, c_S(k' + k_Z) + k'\, k_Z\, c_Z - T\, k_Z(k' + k_Z)}{T\, k_Z}. \qquad (14)$$

The function must therefore necessarily have a minimum which precedes this sharp upturn (4, p. 136). The theory predicts, in short, that for a finite (though possibly quite large) value of the initial stimulus, no finite stimulus increment, however large, can be discriminated. From equation (14) it is seen that this critical value of k varies inversely as the threshold T, thus exhibiting the converse of the relation shown in equation (12).

The existence of such threshold effects as the above seems reasonable enough in view of the known facts. It is by no means easy, however, to see what feature of the mechanism is responsible for this result. A close study of the deductive steps involved in arriving at such predictions suggests that they are connected with the fact that c_B, as given by equation (3), approaches a finite limiting value with increasing k, but never is capable of an infinite range of values. It is illuminating to compare the present model with such a situation as is pre-

sented by molecules crossing a potential energy barrier, where only those molecules with sufficiently high kinetic energies are able to cross a barrier of a given magnitude. Since the usual kinetic energy distribution functions, such as the Maxwellian distribution, admit kinetic energy values approaching infinity, there are always some molecules present which can cross any potential barrier no matter how high; the number crossing drops to zero only as the barrier becomes infinitely high. But when the quantity involved in a threshold never exceeds finite magnitudes, one would expect that the threshold would block completely even when it is only of finite height.

One further point may be worth mentioning. If k or T were increased beyond the critical values given by equations (12) or (14), we would enter a region in which δk is negative. It is not clear to the authors whether to attribute any particular significance to this circumstance, or to regard this region as being psychologically meaningless. Literally interpreted, it would mean that in this region a decrease of stimulus intensity can be discriminated, but not an increase.

In this connection it should be noted that we have previously failed to specify the sign of T, and have in fact implicitly treated it as a positive quantity. However, it is just as reasonable to suppose that a decrease of c_B should correspond to a perception of decreasing the stimulus intensity; in that case we would have to regard T as being positive whenever the organism supposes itself to see an increase in the stimulus, and negative whenever it supposes itself to perceive a decrease in the stimulus.

In that case, the possibility exists that T and δk may be of opposite sign. This is perhaps more obvious if we solve equation (11) for T, giving

$$T = k_0\,\delta k\,\frac{k_0\,c_S(k' + k_Z) + k'\,k_Z c_Z}{\{k_0(k' + k_Z) + k_Z\,k\}\{k_0(k' + k_Z) + k_Z\,k + k_Z\,\delta\,k\}}. \tag{15}$$

Here, if δk is negative, T also is negative until δk reaches such a value that

$$|\delta k| > \frac{k_0(k' + k_Z) + k_Z\,k}{k_Z}. \tag{16}$$

Then T is once more positive. It is tempting to attempt to give an interpretation to such reversals of sign; but any interpretation one can think of seems rather implausible. It is true that errors occur in discrimination experiments, such that an increase of intensity is sometimes perceived as a decrease (4, p. 131). But such errors do not ap-

pear to be sharply divided into zones depending upon the magnitude of the increment of stimulus intensity. On the contrary, they appear to occur with some kind of statistical distribution at almost all intensity levels [see for example the data on the Müller-Lyer illusion in (4)]. It seems likely that such errors of discrimination are concerned with mechanisms of a statistical character, which may be superimposed upon such a fundamental mechanism as our present theoretical model represents, but entirely distinct from this mechanism. It therefore seems advisable to regard these regions of reversal of sign as meaningless by-products of the formal relations (a situation by no means uncommon).

A mechanism closely related to the one discussed here was earlier developed by Hecht, (3, 5). This dealt with the specific field of visual discrimination, and so may be regarded as a special case of our general considerations regarding the steady state. Hecht was of course able to interpret the substances of his theory concretely in terms of the photosensitive substances of the retina and their precursors and breakdown products. His expressions were in another sense somewhat more general than ours, since he assumed one of the chemical reactions to be bimolecular. He did not, however, include a source and sink in his system. Thus his steady state is an equilibrium state, rather than a general steady state in which a non-vanishing flow of matter and energy is involved.

It is obvious how one could examine the effect of the other parameters of the system on c_B, if one were inclined to suspect any one of them as a more likely representative than k for the stimulus intensity. This development may be left aside for the present. We may note in passing, however, that a change of either c_S or c_Z can be ruled out immediately, since either one would produce a response increment simply proportional to the stimulus increment, and independent of the initial stimulus level.

It should be emphasized that we are not proposing the foregoing simple model as a "theory" of psychophysical discrimination processes, but simply as an illustration of what can be done with such an approach. It would be premature to attempt an interpretation of the model by arguing that our substances A and B might be identified with acetylcholine, potassium ion, or the like. Moreover, if one of these substances or any other should prove to be fundamental in the functioning of the central nervous system, it is almost certain that such a function would have to be represented by a more complicated reaction system than the one studied here.

REFERENCES

1. Burton, Alan C. The properties of the steady state compared to those of equilibrium as shown in characteristic biological behavior. *J. cellular and comparative Physiology,* 1939, **14**, 327-349.
2. Taylor, Hugh S. (ed.) A treatise on physical chemistry. New York: D. Van Nostrand Co., Inc., 1931.
3. Hecht, Selig. The visual discrimination of intensity and the Weber-Fechner law. *J. gen. Physiology,* 1924, **7**, 235-267.
4. Guilford, J. P. Psychometric methods. New York: McGraw-Hill Book Co., Inc., 1936.
5. Hecht, Selig. Intensity discrimination and the stationary state. *J. gen. Physiology,* 1927, **6**, 355-373.

PSYCHOMETRIKA—VOL. 10, NO. 1
MARCH, 1945

GROUP DIFFERENCES IN SIZE ESTIMATION

H. L. ANSBACHER*
BROWN UNIVERSITY, PROVIDENCE, R. I.

K. MATHER†
JOHN INNES HORTICULTURAL INSTITUTION, MERTON, LONDON

Fifty-two subjects differing in sex, age, education and domicile (rural or urban) were given the problem of judging the height of an upright board in a natural setting. A preliminary analysis was made on the basis of the simple initial ratio method, both for the original data in feet and for original data converted to log units. Because the effects of interaction of the several variables made the results of this method inconclusive, the analysis of variance technique, as described by Yates(11) for data where the classes are not equally represented, was applied. This technique showed that, while together the four factors markedly affected judgment, sex had no significant individual effect, age had the biggest individual effect but possibly a spurious one, education and domicile had suspiciously large individual effects, and the effect of the four factors may be regarded as simply additive. The relation of the findings to those of previous investigators is discussed. The authors regard as an important result of the analysis the guidance it offers in the design of further experiments, since it demonstrates the value of equal representation for all classes into which data are to be segregated.

I. Problem

In the most exhaustive study on size constancy, the one by Holaday (4), it was found that with the normal (objective) attitude and binocular vision, an object is usually slightly underrated as to its 'real' size. Specifically, constancy judgments of 8-cm. cubes at distances up to 8 m. result in a constancy ratio of 82.3. This means that on a scale of 100, with projective size at the zero end, and actual size at the high ends, the cubes were seen as 17.7 shorter than the actual size. Not expressed in terms of the constancy ratio, but rather in percentages of actual size, the underestimation would in any event be less than 17.7. How much less would depend on the value of the projective size used as the basis of the constancy-ratio calculation. But this information is not given in the Holaday study. His subjects were 10 graduate students in psychology. The study, like all others on size constancy, was conducted in a laboratory setting.

* Responsible for the experiment and general interpretation.
† Responsible for the statistical analysis.

It was the original purpose of the present study to find out how the results of size estimation in a natural setting would compare with Holaday's findings.

II. Method

1. *Procedure.* The important difference in procedure of the present study and the laboratory constancy experiments is that in the latter regular psychophysical procedure (method of limits or of constant stimuli) with a series of comparison objects is used, whereas in the present study we had to resort to purely verbal judgment.

The task was the judgment of the height of an orange-colored upright board, which was 7 feet high and 8 inches wide. Mounted on a tripod, the lower end of the board rose 2 feet above the ground, so that the entire board was visible, unobstructed by grass. It was erected at a distance of 1000 feet from the observer, in a landscape of grass and shrubbery with medium tall trees in the background, about 1500 feet from the point of observation. The erection of a series of comparison objects which would have had to reach a considerable height was not possible. Thus our results are not strictly comparable to those obtained by the usual constancy procedure. Yet we would doubt that the differences found are altogether due to the difference in procedure. Extremely high judgments were verified as follows: The observer's attention was called to the barn in front of which he stood, and he was asked to compare the stimulus object to the corner height of the barn which was 20 feet. The one observer who estimated the stimulus as being 30 feet said, "It is about half as high again as the barn." Others, according to their judgments, said it was about as high as the barn or lower. This way of verification resembles constancy procedure.

The stimulus was hidden from the observer until he had reached the point of observation behind the barn. Then he was asked: "Do you see that orange thing over there? How tall does it look to you?" The response was invariably in even feet.

2. *The Sample.* The study was conducted at the country place of one of the authors in a remote part of Columbia County, New York. For several weeks, any person who happened to pass by and had a few minutes to spare was asked to step to the observation point and render his judgment. In this way a sample of 52 was gathered which differed objectively in 4 factors: Regarding *sex,* 26 of the observers were women and 26 men. In *age* they ranged from 14 to well over 60. *Education* varied from Ph.D. and M.D. degrees down to virtual illiteracy. As to *domicile,* 31 were urban and 21 rural. The urban group, without exception, lived in the New York metropolitan area

and had come to the country for a short visit. Domicile was largely related to occupation and socio-economic status. Of the 9 urban men, 4 were professional and 4 upper-middle-class business people. The 22 urban women were all upper-middle-class or middle-class; 8 were or had been professional women and 6 business women. Of the 17 rural men, 6 were tradespeople (carpenter, roofer, baker, electrician), 9 were farmers and road construction workers, one was a youth of 14, while only one had a white-collar occupation. The 4 rural women were young farm girls or farmers' wives.

Only inspection of the results showed that any one of these 4 factors may have had something to do with the judgments. Thus no personal data were recorded at the time of the experiment, except the name. However, all observers were personally known to the experimenter so that the needed information could be filled in subsequently. Sex and domicile were obvious dichotomies. Accordingly, educational status also was considered to divide the sample into two groups, depending on whether they had been to college or not. The former will hence be referred to as "more educated," the latter as "less educated." For age, the dividing line was drawn from inspection of the data and at an estimated age of 48. Those below 48 will hence be referred to as "young," the others as "old." On this basis the sample would consist of 16 classes, but was actually represented in only 10 classes and with unequal frequencies as shown in Table 3.

III. Results

Judgments at the upper end were spaced at wider intervals than at the lower end, which is in accordance with the Weber-Fechner law. Therefore the general practice with this type of data was followed and each score converted into an arbitrary log unit. This was done from the readings on semi-logarithmic graph paper. When calculations were made on this basis, the values were reconverted into feet where this was called for.

1. *The sample as a whole.* Distribution of the judgments by all 52 observers is shown in Table 1, in feet and log units. The means as obtained from the two scales are given, and that in log units is reconverted into feet.

The mean judgment of 8.6 feet (reconverted) for the 7 ft. board represents an overestimation of 23%, which compares with an underestimation of at least 17% in Holaday's study. However, his sample was limited to students. If we calculate from our sample the mean

TABLE 1
Distribution of Judgments, in Feet and Log Units

Judgments feet	Judgments log units	Frequencies	
3	1	2	
4	5	2	
5	8	5	
6	11	9	$M_{feet} = 9.9$
7	13	2	
8	15	5	$M_{log} = 16.3$
9	17	2	
10	18	7	$M_{feet} = 8.6$
11	20	1	(reconverted)
12	21	7	
15	24	4	
18	27	2	
20	28	3	
30	34	1	
		$n = 52$	

for those 17 individuals who would be most comparable to the students, i.e., the 10 young urban more educated women, the 6 young urban more educated men and the 1 young rural more educated man (see Table 3), we obtain a mean of 5.9 feet (reconverted). This represents an underestimation of almost 16%, a figure coming close to Holaday's. Yet in the light of the discussion below, it would seem that this similarity is purely a coincidence.

2. *Preliminary analyses.** Four factors were recognized: sex, age, education and domicile, each of which divides the observers into two groups. The mean judgments, both in feet and log units, for each of the two groups obtained by the use of each factor are shown in Table 2. Lower judgments are given by the women, the young, the more educated and the urban. If, however, the differences obtained from these classifications are tested by comparison with their standard errors, it is found that they are not all equally significant. The quantity t given in Table 2 is the ratio of the difference to its standard error. It may be found either directly in this way or by a simple analysis of variance (see 7). Each t has 50 degrees of freedom, since out of the 51 degrees of freedom between 52 observers, 1 is used up in making the comparison between men and women, or between old and young etc.

* The statistical notation used throughout follows that of Fisher (2) and Mather (7). In particular, S is used to indicate summation, a process which is often shown by Σ.

TABLE 2
Simple Tests of Effects of the Four Factors on Judgment

Factor	Class	Number in Class	In Feet				In Log Units			
			Mean	Difference	t	Probability	Mean	Difference	t	Probability
Sex	Women	26	8.9	2.1	1.452	0.2 — 0.1	14.6	3.4	1.679	.01
	Men	26	11.0				18.0			
Age	Young	39	8.6	5.4	3.543	0.001	14.5	7.2	3.374	0.001
	Old	13	14.0				21.7			
Education	More Educated	17	6.4	5.2	3.635	0.001	10.8	8.2	4.433	very small
	Less Educated	35	11.6				19.0			
Domicile	Urban	31	8.9	3.7	3.296	0.001	13.7	6.3	4.508	very small
	Rural	21	12.6				20.0			

Whether the calculation is made using the initial judgments in feet, or the converted judgments in log units, t has a probability of somewhere near 0.1 for the men-women difference, which is thus not significant on either test. For each of the other three factors, on the other hand, t has a probability of 0.001 or lower, using either feet or

log units, and so each of the three factors might be considered to be exercising a real effect on the judgment of height.

This simple test cannot, however, be regarded as trustworthy, for the following reason. When all four classificatory factors are used, we can simultaneously distinguish 16 classes, young more educated urban women, young more educated urban men, young more educated rural women and so on. It will be seen from Table 3 that

TABLE 3

Composition of the Sample and Mean Judgments of the Classes

Sex→		Women						Men					
Education→		More Educated			Less Educated			More Educated			Less Educated		
			Mean			Mean			Mean			Mean	
Age	Domicile	Number	Feet	Logs	Number	Feet	Logs	Number	Feet	Logs	Number	Feet	Logs
Young	Urban	10	5.6	8.7	5	9.8	17.0	6	7.2	13.0	3	6.7	12.0
	Rural		—		2	6.0	10.0	1	10.0	18.0	12	12.0	20.1
Old	Urban		—		7	11.9	20.0		—			—	
	Rural		—		2	15.5	24.0		—		4	17.0	23.5

TABLE 4
Rank Orders of Scores and Mean Judgments

Class	Score	Total Index	Mean Judgment	Order of Total Indices	Order of Means
Female young more educated urban	1.1.1.1	4	5.2	1	1
" " less educated "	1.1.2.1	5	9.0	2.5	5
" " " rural	1.1.2.2	6	5.6	5.5	2
" old " urban	1.2.2.1	6	11.0	5.5	7.5
" " " rural	1.2.2.2	7	15.0	8.5	10
Male young more educated urban	2.1.1.1	5	7.0	2.5	4
" " " rural	2.1.1.2	6	10.0	5.5	6
" " less educated urban	2.1.2.1	6	6.6	5.5	3
" " " rural	2.1.2.2	7	11.0	8.5	7.5
" old " "	2.2.2.2	8	14.5	10	9

Rank order correlation is .80

only 10 of these 16 classes are represented by actual observers in the data, and these 10 classes are not equally filled, the numbers of observers falling into them varying from 1 to 12. Thus, in taking, say, the comparison of rural as opposed to urban, the result will not be independent of the other three factors. There are, for example, 20 less educated rurals and 16 more educated urbans, against 15 less educated urbans and 1 more educated rural.

Furthermore, it can easily be shown by the calculation of a rank correlation that the factors supplement one another. Each of the 10 classes was given a score for each factor. The score of 1 was given when the class contained the observers giving the lower judgment on the classification in question, and correspondingly a score of 2 was given to the higher judging group. Thus women, young, more educated and urban all take scores of 1, while men, old, less educated and rural all take scores of 2. A female, old, more educated and urban would, for example, have a score of 1.2.1.1. The scores for each of the 10 classes are given in Table 4. The four digits in each score are then added to give a total index which can be correlated with the mean judgment (calculated from log units and reconverted to feet) of the class. The rank order correlation obtained in this way is .80 and shows well how the effects of the four factors supplement one another. Thus our simple tests of the significance of effect of each of the four factors must be misleading. To take the example quoted above, there are a greater number of rural less educated and urban more educated than there are of rural more educated and urban less educated. The simple difference between urban and rural must thus include a supplementary reinforcing contribution from the more educated versus less educated classification. A more rigorous statistical analysis is necessary.

3. *Full Analysis.* The method of analyzing data of the present type, where the classes are not all equally represented, some even being entirely missing from the multiple classification, is discussed by Yates (11). He describes a method of fitting constants, to represent the effects of each classificatory factor, by means of the least squares technique.

As applied to our data, we propose in effect, on this method, to find a best fitting constant b_s which may be taken to represent the effect of sex on judgment, a second b_a to represent the effect of age, b_e and b_d to represent education and domicile effects. These constants are calculated so that the summed squares of the differences between observed judgments and those expected on the basis of the four constants are at a minimum. The analysis is formally equivalent to that of a multiple regression analysis, fully described by Fisher (2) and

Mather (7). Those accounts will give all the details of procedure.

Let us first analyze the data in the form of the initial judgment in feet. The first step is, as in calculating the rank correlation, to assign scores representing the classifications. These may be arbitrary since each factor divides the data into only two classes, and hence -1 and 1 have been used in each case for the calculations. (The corresponding scores of 1 and 2 as used in Table 4, or indeed any other scores, however, could have been taken without affecting our ultimate tests of significance.)

Where

x_s is the sex score, $x_s = -1$ for women and $x_s = 1$ for men: similarly

x_a is the age score, $x_a = -1$ for young and $x_a = 1$ for old,

x_e is the education score, $x_e = -1$ for more educated and $x_e = 1$ for less educated,

x_d is the domicile score, $x_d = -1$ for urban and $x_d = 1$ for rural.

When y is the observed judgment in feet and Y the corresponding expectation derived from the constants, we can represent our problem as that of calculating the regression coefficients $b_s \cdots b_d$ in the equation

$$Y = \bar{y} + b_s(x_s - \bar{x}_s) + b_a(x_a - \bar{x}_a) + b_e(x_e - \bar{x}_e) + b_d(x_d - \bar{x}_d)$$

such that $S(y - Y)^2$ is a minimum.

The appropriate values of $b_s \cdots b_d$ are given by the solutions of the four equations:

$$\begin{aligned} b_s S(x_s - \bar{x}_s)^2 + b_a S[(x_a - \bar{x}_a)(x_s - \bar{x}_s)] + b_e S[(x_e - \bar{x}_e)(x_s - \bar{x}_s)] \\ + b_d S[(x_d - \bar{x}_d)(x_s - \bar{x}_s)] = S[(x_s - \bar{x}_s)(y - \bar{y})] \end{aligned}$$

$$\begin{aligned} b_s S[(x_s - \bar{x}_s)(x_a - \bar{x}_a)] + b_a S(x_a - \bar{x}_a)^2 + b_e S[(x_e - \bar{x}_e)(x_a - \bar{x}_a)] \\ + b_d S[(x_d - \bar{x}_d)(x_a - \bar{x}_a)] = S[(x_a - \bar{x}_a)(y - \bar{y})] \end{aligned}$$

$$\begin{aligned} b_s S[(x_s - \bar{x}_s)(x_e - \bar{x}_e)] + b_a S[(x_a - \bar{x}_a)(x_e - \bar{x}_e)] \\ + b_e S(x_e - \bar{x}_e)^2 + b_d S[(x_d - \bar{x}_d)(x_e - \bar{x}_e)] \\ = S[(x_e - \bar{x}_e)(y - \bar{y})] \end{aligned}$$

$$\begin{aligned} b_s S[(x_s - \bar{x}_s)(x_d - \bar{x}_d)] + b_a S[(x_a - \bar{x}_a)(x_d - \bar{x}_d)] \\ + b_e S[(x_e - \bar{x}_e)(x_d - \bar{x}_d)] + b_d S(x_d - \bar{x}_d)^2 \\ = S[(x_d - \bar{x}_d)(y - \bar{y})]. \end{aligned}$$

The items of the types $S(x_s - \bar{x}_s)^2$ and $S[(x_a - \bar{x}_a)(x_s - \bar{x}_s)]$ are found from the scores described above. Now $S(x_s - \bar{x}_s)^2 = S(x_s^2) - \dfrac{S^2(x)}{n}$ where n is the number of observers. There are 26 observers

(women) with a score of $x_s = -1$ and 26 (men) with $x_s = 1$. Thus $S(x_s^2) = 26 \times (-1)^2 + 26 \times 1^2 = 52$ and $\frac{1}{n} S^2(x) = \frac{1}{52} [26 \times (-1) + 26 \times 1] = 0$.

So $S(x_s - \bar{x}_s)^2 = 52 - 0 = 52$.

Similarly $S[(x_a - \bar{x}_a)(x_s - \bar{x}_s)] = S(x_a x_s) - \dfrac{S(x_a)S(x_s)}{n}$.

We have 17 young females with $x_s = -1$ and $x_a = -1$, 22 young males with $x_s = 1$ and $x_a = -1$, 9 old females with $x_s = -1$ and $x_a = 1$, and 4 old males with $x_s = 1$ and $x_a = 1$.
Then

$$\begin{aligned} S(x_a x_s) &= [17 \times (-1) \times (-1)] + [22 \times 1 \times (-1)] \\ &\quad + [9 \times (-1) \times 1] + [4 \times 1 \times 1] = 0 . \end{aligned}$$

$$S(x_s) = 0 \text{ and } S(x_a) = -26 \text{ giving } \frac{1}{n} S(x_a) S(x_s) = 0.$$

Thus

$$S[(x_a - \bar{x}_a)(x_s - \bar{x}_s)] = -10 - 0 = -10.$$

The remaining sums of squares and sums of cross products, including those involving y, the judgment in feet, are found in the same way. The four equations may then be written down. It is, however, preferable to solve the following four sets each of four equations, the left sides of which are the same as those given above, but the right sides of which have 1,0,0,0; 0,1,0,0; 0,0,1,0; and 0,0,0,1 substituted for $S[(x_s - \bar{x}_s)(y - \bar{y})]$ etc. The solutions of these latter equations are necessary for the calculation of the standard errors and also aid other calculations.

We thus have the following equations for solution:

$$\begin{aligned} 52.00b_s - 10.00b_a + 6.00b_e + 26.00b_d &= 1,0,0,0 \\ -10.00b_s + 39.00b_a + 17.00b_e + 3.00b_d &= 0,1,0,0 \\ 6.00b_s + 17.00b_a + 45.77b_e + 19.46b_d &= 0,0,1,0 \\ 26.00b_s + 3.00b_a + 19.46b_e + 50.08b_d &= 0,0,0,1 \end{aligned}$$

and the solution gives us 16 values generally denoted as

$$\begin{matrix} C_{11} & C_{21} & C_{31} & C_{41} \\ C_{12} & C_{22} & C_{32} & C_{42} \\ C_{13} & C_{23} & C_{33} & C_{43} \\ C_{14} & C_{24} & C_{34} & C_{44} \end{matrix}$$

etc., where C_{11} is the value obtained for b_s from the first set of equations, C_{34} that for b_e from the fourth set and so on. This C matrix of 16 solutions turns out to be

0.02853	0.00871	−0.00055	−0.01512
0.00871	0.03365	−0.01301	−0.00148
−0.00055	−0.01301	0.03151	−0.01118
−0.01512	−0.00148	−0.01118	0.03225

Now

$$\begin{aligned} b_s &= C_{11}\,S[(x_s-\bar{x}_s)(y-\bar{y})] + C_{12}\,S[(x_a-\bar{x}_a)(y-\bar{y})] \\ &\quad + C_{13}\,S[(x_e-\bar{x}_e)(y-\bar{y})] + C_{14}\,S[(x_d-\bar{x}_d)(y-\bar{y})] \\ &= (0.02853\times 54) + (0.00871\times 106) - (0.00055\times 119.38) \\ &\quad - (0.01512\times 113.23) \\ &= 0.6862 \\ b_a &= C_{21}\,S[(x_s-\bar{x}_s)(y-\bar{y})] + C_{22}\,S[(x_a-\bar{x}_a)(y-\bar{y})] \\ &\quad + C_{23}\,S[(x_e-\bar{x}_e)(y-\bar{y})] + C_{24}\,S[(x_d-\bar{x}_d)(y-\bar{y})] \\ &= 2.3165 \\ b_e &= 1.0870 \\ b_d &= 1.3436\,. \end{aligned}$$

The sum of squares, $S(y-\bar{y})^2$, between the 10 classes is found to be 635.6352 of which an item

$$\begin{aligned} &b_s\,S[(x_s-\bar{x}_s)(y-\bar{y})] + b_a\,S[(x_a-\bar{x}_a)(y-\bar{y})] \\ &+ b_e\,S[(x_e-\bar{x}_e)(y-\bar{y})] + b_d\,S[(x_d-\bar{x}_a)(y-\bar{y})], \end{aligned}$$

or 564.5057, is ascribable to the action of the four factors: sex, age, education and domicile. This leaves 71.1295 as residual variation due to difference between y and Y, i.e., $S(y-Y)^2$. Since there are 9 degrees of freedom between the 10 classes, 4 of which are taken up in calculating $b_s \cdots b_d$, this residual sum of squares corresponds to 5 degrees of freedom.

There is, however, a further residual sum of squares, viz., that from variation in judgment between members of the same class. This can be found by calculating the sum of squares between all 52 judgments, which turns out to be 1435.6923, and subtracting from it the item, 635.6352, found as the sum of squares between classes. The remainder is the sum of squares within classes. We thus obtain the analysis of variance of judgment in feet as shown in Table 5.

The allocation of the degrees of freedom is not difficult to see. There are 51 in all, between 52 observers, of which 9 are between the 10 class means. The sum of squares within classes thus corresponds to 42 degrees of freedom.

The mean squares are found as the ratio of the corresponding sums of squares to their numbers of degrees of freedom. Taking the mean square within classes as the estimate of error variation, the other two mean squares, that ascribable to the four classificatory factors and the residuum between classes after fitting constants repre-

TABLE 5
Analysis of Variance of Data in Feet

Item	Sum of Squares	Degrees of Freedom	Mean Square	Variance Ratio	Probability
Factors: Main Effects	564.5057	4	141.1264	7.41	very small
Factors: Interaction	71.1295	5	14.2259		not significant
Within classes (first error)	800.0571	42	19.0490		
Total	1435.6923	51			
Pooled Error	871.1866	47	18.5359		

senting these factors, may be tested for significance by comparison with it. The residuum between classes is clearly not significant as its mean square is somewhat less than that for error. The other mean square, for differences ascribable to the classificatory factors, is next divided by the error mean square to give a variance ratio. The variance ratio should be entered in a table of variance ratios such as those provided by Fisher and Yates (3), or Mather (7). On entering in such a table, the item ascribable to the main effects of the four factors is found to have a very significantly low probability. The four factors, taken together, are having an effect on judgment. But since the residuum, which tests for "interaction" of the four factors, i.e., for departure from additiveness of their effects, is insignificant, they may be regarded as additive in their effects.

As this residual or "interaction" item is not significant, it may be pooled with the item for variation within classes to give a pooled estimate of error based on 47 degrees of freedom, the mean square being 18.5359.

Now the standard error of b_s is found by multiplying $\sqrt{C_{11}}$ by the square root of the error mean square and so is $\sqrt{18.5359} \times \sqrt{0.02853} = \pm\ 0.7272$. The standard errors of b_a, b_e and b_d are similarly found, by using $\sqrt{C_{22}}$, $\sqrt{C_{33}}$ and $\sqrt{C_{44}}$ in place of $\sqrt{C_{11}}$, to be 0.7898, 0.7642, and 0.7731, respectively. Thus b_s is less than its standard error, while $t_a = \dfrac{b_a}{\text{standard error of } b_a} = 2.933$, $t_e = 1.422$, and $t_d = 1.738$. Each t has 47 degrees of freedom, since the error mean square was found from this number of independent comparisons, and entry in a table of t gives the probability (P) as

$$\begin{aligned} &t_s\,, P = 0.4 - 0.3 \\ &t_a\,, P = 0.01 - 0.001 \\ &t_e\,, P = 0.2 - 0.1 \\ &t_d\,, P = 0.1 - 0.05\,. \end{aligned}$$

Thus the effect ascribable unambiguously to age is clearly significant, that for sex clearly insignificant, and those for education and domicile of doubtful significance.

It may seem remarkable that, while the four factors taken together have such a markedly significant effect, as shown by the analysis of Table 5, only one of them, age, has a significant effect taken by itself. The explanation is that t_s, t_a, etc., test the effects *unambiguously* ascribable to sex, age, etc. There remain, however, significant effects not ascribable unambiguously to any one factor, owing to the unequal representation of the 16 classes, but which may be traced

to the four factors acting together. In the present case this item can be shown to represent nearly half of the sum of squares attributed to the four factors in Table 5. In other words, the design of the experiment is such that only half the information available can be interpreted; the other half cannot be used to tell us anything about the actions of the four factors as individuals. Only by having all 16 classes represented by equal numbers of observers can this loss of information be wholly prevented. It may be added, too, that in such a case the analysis would be much less laborious than that which it was necessary to use for the present unbalanced data.

The conclusions to which the analysis of the data in feet have led are changed in no material particular when the converted data in log units are used. Since the scores $x_s \cdots x_d$ are assigned independently of the actual judgments, the analysis as conducted so far as the C matrix is the same for log units as for feet. In log units

$$\begin{aligned} S[(x_s - \bar{x}_s)(y - \bar{y})] &= 87.00 \\ S[(x_a - \bar{x}_a)(y - \bar{y})] &= 140.50 \\ S[(x_e - \bar{x}_e)(y - \bar{y})] &= 187.81 \\ S[(x_d - \bar{x}_d)(y - \bar{y})] &= 157.88 \,. \end{aligned}$$

We find

$$\begin{aligned} b_s &= 1.2154 \pm 0.9765 & \qquad t_s &= 1.245\,, \; P = 0.3\text{-}0.2 \\ b_a &= 2.8085 \pm 1.0600 & t_a &= 2.650\,, \; P = 0.02\text{-}0.01 \\ & & \text{and} & \\ b_e &= 2.2770 \pm 1.0262 & t_e &= 2.219\,, \; P = 0.05\text{-}0.02 \\ b_d &= 1.4685 \pm 1.0382 & t_d &= 1.414\,, \; P = 0.20\text{-}0.10\,. \end{aligned}$$

The analysis of variance becomes that given in Table 6, the calculations being made exactly as before but with the log data. Again we find the interaction mean square to be insignificant (though now slightly greater than error). It is pooled with the mean square within classes to give a pooled estimate of error for 47 degrees of freedom, from which the standard errors of the four b constants are calculated. As before, the sex effect is insignificant, though the age effect is significant. Education effects seem to be clearly significant now, though they were not clearly so in terms of the other units, while domicile is rather less significant than before.

Again as before, the item testing the pooled effects of all four factors, in the analysis of variance, is much more significant than any of the four individual tests would suggest. This time, however, the information which is uninterpretable in terms of one or other of

TABLE 6
Analysis of Variance of Data in Log Units

Item	Sum of Squares	Degrees of Freedom	Mean Square	Variance Ratio	Probability
Factors: Main Effects	1159.8242	4	289.9561	8.99	very small
Factors: Interaction	216.8322	5	43.3664		not significant
Within classes (first error)	1354.0167	42	32.2385		
Total	2730.6731	51			
Pooled Error	1570.8489	47	33.4223		

the four individual effects is somewhat more than half that contained in the items of the analysis of variance.

We may thus conclude that the four factors taken together markedly affect judgment of height, though owing to the nature of the data, only age can be shown to have a marked individual effect. Since the dividing line between "young" and "old" was taken to be 48 as a result of inspection of the data themselves, the significance of even this factor must be viewed with some suspicion. Sex has no significant effect, and education and domicile have suggestive, though not unambiguously significant, individual effects. The effects of the four factors may be regarded as simply additive.

The most important result of the analysis is, however, the guidance that is given us in designing further experiments. It is clear that the unequal representation of observers in the 16 classes has resulted in much of the information being uninterpretable in terms of the four individual effects, to say nothing of its leading to a more laborious statistical analysis. It is thus clear that in the future equal representation should be aimed at, even if it involves some extra trouble, for only in this way can the full value of the experimental procedure be realized.

IV. Discussion

As stated above, age as the strongest factor may be partly an artifact because here classification was based on inspection of the data alone. Furthermore, while the other two effective factors are environmentally determined, age is a physiological factor, and it is not possible here to go into the physiology of perception. Sex, on the other hand, was found to have no real effect. Therefore we shall limit the discussion to the suggestive, though not unambiguously significant, effects of education and domicile.

1. *Relationship to intelligence* of the results from the present study and those from constancy experiments. The suspiciously effective factors of education and domicile—and incidentally age as well—have in common that they correlate with scores on intelligence tests, as has been shown in numerous studies. In addition, education and domicile in our sample coincided largely with occupation, a further factor known to be related to test intelligence. Thus, judgment of size would be related to intelligence, lower judgments being found with higher intelligence, i.e., educated, urban subjects.

In several laboratory constancy experiments also, those with higher test intelligence, or the educated European-Americans, were found to perceive things as smaller than the less intelligent, or other

ethnic groups. These experiments are: (a) Thouless (9) with 53 subjects found a correlation of $-.41 \pm 0.01$ between intelligence, as measured by the Cattell test and the NIIP, and size and shape constancy; (b) Klimpfinger (5) found a size constancy ratio of 95.8 for 20 students, which compares to a ratio of 99.7 found previously by another investigator for rural hired men and women; (c) Thouless (10) in another study, found a size constancy ratio of 0.61 for 49 English students and one of 0.76 for 20 Indian students, with a critical ratio of 4.3; (d) Beveridge (1) found a ratio of 0.75 for 8 Europeans and one of 0.88 for 44 West African natives, students of drawing at a Presbyterian training college: (e) Sheehan (8) using 25 young college women found a correlation of -0.375 between size constancy and scores on the Street *Gestalt Completion Test* which "measures a non-verbal aspect of mental organization."

While our findings agree with the constancy results regarding the inverse relationship between *height* of judgment and intelligence, they disagree regarding *accuracy* of judgment, which in our experiment was positively related to intelligence. The following is an attempt to explain this discrepancy.

2. *Effect of the experimental situation.* In the constancy experiments, the individual who makes a high score (great accuracy) is said to have an objective attitude, that of naive observation, and the one who makes a low score (less accuracy) is said to have a subjective or analytical attitude, the latter being found with greater intelligence.

Our findings and observations would confirm the existence of these two attitudes and their effect on accuracy of judgment, but not their dependency on intelligence, nor their effect on height of judgment. Both these relationships are determined by the experimental situation, as follows:

In the constancy experiments it is the less intelligent who assume the naive attitude because a laboratory situation is unfamiliar, means nothing to the primitive, the child, the less intelligent; it is not part of his world. Thus he yields to the impression without further reflection, always the better way to respond to tasks of this kind. To the more intelligent, however, it may be an intellectual challenge, resulting in the analytical attitude with consequent greater deviation from the actual size.

Our situation, on the other hand, offered a greater challenge to the less intelligent. The strange thing in the landscape meant nothing to the more educated professional city people who thus yielded to the impression without much reflection, in the naive attitude. They approached the task with a certain indifference. But to the less educated

rural people the task was a challenge which they took up with interest and confidence in their experience. This resulted in the analytical, subjective attitude. To judge the size of an object in the landscape was part of their practical, everyday lives; the object might be something in relation to their work. Thus their level of aspiration was high, whereas that of the city people was low. Aside from general observations, these considerations are supported by one extreme case. The lowest judgment among the rural people by far was made by a happy-go-lucky, unsettled young woman of the lowest rural socio-economic level. Being completely indifferent, she was not challenged by the task.

The experimental situation also determines the direction of the error. In our experiment the 7-ft. board in its setting could not well be underestimated more than 3-4 feet, whereas there was no limit to overestimation. Thus those who tended to larger error found their outlet on the upper side. In the constancy experiments the perceptual laws underlying the constancy phenomenon and the experimental procedure favor the error of underestimation.

The discrepancy stated at the end of section IV, 1 finds its resolution then in the realization that intelligence is not the decisive factor in the results of these two types of experiment, but rather attitude. Attitude will be determined by the experimental situation, which also determines the direction of the error.

3. *Theory.* On the basis of numerous animal experiments and the human constancy experiments cited above, Locke (6) proposed as a general theory, to be valid within the human species as well, that perceptual ability decreases as a given phylum becomes more complex and its intelligence increases. This would be because intelligence displaces perception as the principal means of coping with the environment.

Our considerations would disclaim the validity of this theory for the human species. The apparent "perceptual ability" of the less intelligent in the human constancy experiments is probably only the result of the laboratory situation if in an actual outdoor situation the more intelligent show so much more "perceptual ability." Furthermore, Locke's theory involves the assumption that perceptual constancy is a unitary factor. But this assumption was not verified in a comprehensive study by Sheehan (8) designed to settle this problem.

On the basis of the present results and considerations we would conclude that at the human level perceptual function, like so many others, is determined largely by the individual's background and interests, his social and individual personality, his level of aspiration—but not by his intelligence.

V. Summary and Conclusions

Fifty-two observers, heterogeneous as to sex, age, education and domicile, judged the size of an orange-colored upright board that was 8 inches wide and 7 feet high, erected at a distance of 1000 feet from the observer in a landscape of grass and shrubbery. The following results were found:

1. The sample as a whole overestimated the size by 23% [mean judgment 8.6 feet (reconverted), range 3 to 30 feet].

2. The 17 more educated subjects underestimated the size by 16%, which is similar to size-underestimation of at least 17% by students in a constancy experiment. Yet this similarity appears to be a mere coincidence.

3. Preliminary analysis of the data, whether in feet or transferred into log units, showed that lower judgments were made by the women, the young, the more educated and the urban, as compared with the men, the old, the less educated and the rural people, the sex difference, however, being insignificant.

4. More rigorous analysis showed that, while taken together the four factors markedly affected judgment, (a) sex had no real individual effect, (b) age had the biggest individual effect but possibly a spurious one due to the treatment of the data, (c) education and domicile had suspiciously large individual effects, (d) the four factors were simply additive in effect.

5. In the present experiment higher judgment meant greater error. Thus greater error was found with lower education and rural domicile, both factors known to be related to lower test intelligence. This is in disagreement with (a) the results of certain constancy experiments where greater accuracy was found with lower intelligence and (b) a general theory proposed by Locke (6) on the basis of these constancy results, according to which perceptual efficiency at the human level as well as on the sub-human level is inversely related to intelligence.

6. While thus intelligence cannot be regarded as the factor influencing accuracy of judgment in both our experiment and the constancy experiments, attitude can be regarded as the common factor. The naive attitude would seem to make for greater accuracy of judgment in both situations—it being present among the less intelligent in the constancy experiments and among the more intelligent in our experiment as determined by the experimental situation.

REFERENCES

1. Beveridge, W. M. Racial differences in phenomenal regression. *Brit. J. Psychol.*, 1935, 26, 59-62.
2. Fisher, R. A. Statistical methods for research workers. 6th ed. Edinburgh: Oliver and Boyd. 1936.
3. Fisher, R. A. and Yates, F. Statistical tables for biological, agricultural and medical research. 2nd ed. Edinburgh: Oliver and Boyd. 1943.
4. Holaday, B. E. Die Grossenkonstanz der Sehdinge. *Arch. ges. Psychol.*, 1933, 88, 419-486.
5. Klimpfinger, S. Die Entwicklung der Gestaltkonstanz vom Kind zum Erwachsenen. *Arch. ges. Psychol.*, 1933, 88, 599-628.
6. Locke, N. M. Perception and intelligence: their phylogenetic relation. *Psychol. Rev.*, 1938, 45, 335-345.
7. Mather, K. Statistical analysis in biology. London: Methuen. 1943.
8. Sheehan, M. R. A study of individual consistency in phenomenal constancy. *Arch. Psychol.*, N. Y., 1938, No. 222.
9. Thouless, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, 22, 216-241.
10. Thouless, R. H. A racial difference in perception. *J. soc. Psychol.*, 1933, 4, 330-339.
11. Yates, F. The principles of orthogonality and confounding in replicated experiments. *J. agric. Sci.*, 1933, 23, 108-145.

PSYCHOMETRIKA—VOL. 10, NO. 1
MARCH, 1945

THE RELIABILITY OF COMPONENT SCORES

FREDERICK B. DAVIS
COOPERATIVE TEST SERVICE OF THE AMERICAN COUNCIL ON EDUCATION*

A method is given for determining the reliability of each of the components resulting from a factor analysis by the principal axis method.

Determination of the reliability coefficient of scores in each of the principal components derived from a factorial analysis is desirable for several reasons. It is obvious that the naming and interpretation of components the variances of which cannot be proved greater than might be yielded by chance alone may be misleading and deceptive. Consequently, Hotelling and Thurstone long ago considered this problem, and more recently Hoel† described a precise test for determining the number of significant components that could be obtained by factorial analysis from a given matrix.

A somewhat different approach to determining the statistical significance of components has been suggested by Kelley.‡ First, he has provided a variance-ratio test of the uniqueness of principal-axis components as they exist at any stage of the Kelley iterative process for their determination.§ By means of this test it is possible to ascertain the likelihood that a given component will re-occur in subsequent samples by discovering whether the magnitudes of the variances of successive components are significantly different. Second, he has suggested that whether the variance of a principal component is greater than might be yielded by chance can be determined by noting whether the reliability coefficient of scores in the component is significantly greater than zero. This follows, of course, from the fact that the non-chance variance of a variable is directly proportional to the reliability coefficient of the variable.

It might be supposed that the magnitudes of the reliability coefficients of the principal components obtained from any given matrix

* On leave for military service.

† P. G. Hoel. A significance test for component analysis. *Annals of Math. Stat.*, 8, 149-158.

‡ The writer is indebted to Dr. T. L. Kelley for pointing out the importance of this problem and outlining the method of obtaining component reliabilities.

§ T. L. Kelley. A variance-ratio test of the uniqueness of principal-axis components as they exist at any stage of the Kelley iterative process for their determination. *Psychometrika*, 1944, 9, 199-200.

would be directly proportional to the magnitudes of the component variances, but that is not mathematically necessary and has not proved to be the case in actual practice. The writer has obtained reliability coefficients for the principal-axis components derived from two entirely different matrices. Computation of the coefficients for nine components derived from the first matrix was done by an empirical procedure which is very laborious;* computation of the coefficients for fourteen components derived from the second matrix was done by specific application of the formulas presented later in this article. In both cases, there was the expected tendency for the larger components to have the higher reliability coefficients but the agreement was far from perfect. In the set of nine components, the first, second, third, seventh, and eighth were found to have reliability coefficients sufficiently greater than zero to warrant the belief that their variances were significant. In the set of fourteen components, the reliability coefficients of the initial variables were based on a far larger sample and only the ninth and fourteenth components had reliability coefficients sufficiently low as to warrant discarding the components in interpreting the results of the study.

A test of the sort described by Hoel indicates only the *number* of significant components that may be obtained, while if the reliability coefficient of each component has been computed, the components the variances of which are not significantly greater than zero *may be identified* and discarded. Furthermore, if individual scores in the useful components are to be obtained, the standard error of measurement of obtained component scores may easily be estimated. Since the reliability coefficients of most of the components obtained from a given matrix are likely to be low in comparison with the reliability coefficients of the majority of achievement and aptitude tests, their significance is an important consideration, the determination of which justifies the expenditure of considerable labor. In fact, the writer will go so far as to say that if a factorial analysis is worth doing, it is worth the labor required to obtain the reliability coefficients of the components found.

Let the first component of a matrix containg n initial variables be denoted as

$$C_I = b_1x_1 + b_2x_2 + \cdots + b_nx_n \, , \tag{1}$$

where $b_1 , b_2 , \cdots , b_n$ are regression coefficients derived from a matrix by a principal-axis method, and $x_1 , x_2 , \cdots , x_n$ are scores in the initial variables expressed as deviations from their respective means.

* Frederick B. Davis. Fundamental factors of comprehension in reading. *Psychometrika,* 1944, 9, 185-197.

If we let x_a and x_A represent equivalent halves of x_1, x_b and x_B represent equivalent halves of x_2, etc., then C_a and C_A must represent equivalent halves of C_I and may be denoted as

$$C_a = b_1x_a + b_2x_b + \cdots + b_nx_n\,, \tag{2}$$

$$C_A = b_1x_A + b_2x_B + \cdots + b_nx_N\,. \tag{3}$$

Also,

$$\sigma_1 = \sqrt{\sigma_a^2 + \sigma_A^2 + 2\sigma_a\sigma_A r_{aA}}\,. \tag{4}$$

Since x_a and x_A, x_b and x_B, $\cdots$, x_n and x_N are equivalent halves, equation (4) may be written as

$$\sigma_1 = \sqrt{2\sigma_a^2 + 2\sigma_a^2 r_{aA}}\,. \tag{5}$$

Solving for σ_a, we obtain

$$\sigma_a = \frac{\sigma_1}{\sqrt{2(1 - r_{aA})}}. \tag{6}$$

The fact that x_a and x_A, x_b and x_B, $\cdots$, x_n and x_N are equivalent halves also leads to the fact that all possible intercorrelations of the halves of successive pairs of initial variables are equal to a very close approximation; that is,

$$r_{ab} \equiv r_{aB} \equiv r_{Ab} \equiv r_{AB} \tag{7}$$

$$r_{ac} \equiv r_{aC} \equiv r_{Ac} \equiv r_{AC} \tag{8}$$

$$\begin{matrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{matrix}$$

$$r_{mn} \equiv r_{mN} \equiv r_{Mn} \equiv r_{MN}\,. \tag{9}$$

Since the intercorrelations of the initial variables are known, it is possible to secure numerical values for the intercorrelations of the halves.

$$r_{12} \equiv r_{(x_a+x_A)(x_b+x_B)} = \frac{\sum(x_a + x_A)(x_b + x_B)}{\sqrt{\sum(x_a + x_A)^2}\,\sqrt{\sum(x_b + x_B)^2}}. \tag{10}$$

If equation (10) is expanded and simplified by making use of the identities in (7), (8), and (9),

$$r_{12} = \frac{2r_{ab}}{\sqrt{1 + r_{aA}}\,\sqrt{1 + r_{bB}}}. \tag{11}$$

Solving equation (11) for r_{ab}, we obtain

$$r_{ab} = \frac{r_{12}\sqrt{1 + r_{aA}}\,\sqrt{1 + r_{bB}}}{2}. \tag{12}$$

By means of equations analogous to (12) all the desired intercorrelations of halves of the initial variables may be obtained.

Returning now to equations (2) and (3), the correlation of $r_{C_aC_A}$ is, by definition,

$$r_{C_aC_A} = \frac{\sum C_a C_A}{\sqrt{\sum C_a^2}\sqrt{\sum C_A^2}}. \tag{13}$$

Since C_a and C_A are equivalent, equation (13) may be written

$$r_{C_aC_A} = \frac{\sum C_a C_A}{\sum C_a^2}. \tag{14}$$

Substituting in equation (14), we obtain

$$r_{C_aC_A} = \frac{\sum (b_1x_a + b_2x_b + \cdots + b_nx_n)(b_1x_A + b_2x_B + \cdots + b_nx_N)}{\sum (b_1x_a + b_2x_b + \cdots + b_nx_n)^2}. \tag{15}$$

Expanding and simplifying equation (15), we may write

$$r_{C_aC_A} = \frac{b_1^2\sigma_a^2r_{aA} + b_2^2\sigma_b^2r_{bB} + \cdots + b_n^2\sigma_n^2r_{nN} + 2b_1b_2\sigma_a\sigma_br_{ab} + 2b_1b_3\sigma_a\sigma_cr_{ac} + \cdots + 2b_mb_n\sigma_m\sigma_nr_{mn}}{b_1^2\sigma_a^2 + b_2^2\sigma_b^2 + \cdots + b_n^2\sigma_n^2 + 2b_1b_2\sigma_a\sigma_br_{ab} + 2b_1b_3\sigma_a\sigma_cr_{ac} + \cdots + 2b_mb_n\sigma_m\sigma_nr_{mn}}. \tag{16}$$

With the exception of the terms written r_{aA}, r_{bB}, $\cdots$, r_{nN}, the method of obtaining numerical values for all the terms in equation (16) has been mentioned. The terms r_{aA}, r_{bB}, $\cdots$, r_{nN} are the correlations of one half of each initial variable with the other half. They may best be obtained directly, although they can be estimated by means of the Spearman-Brown formula from the reliability of each entire initial variable.

After a numerical value has been obtained for $r_{C_aC_A}$, the reliability coefficient for component I may be estimated by means of the Spearman-Brown formula.

Using the same correlation coefficients and the same standard deviations of the halves of the initial variables together with the appropriate different sets of b's, one may estimate the reliability coefficient of each other principal component in a similar manner.

PSYCHOMETRIKA—VOL. 10, NO. 1
MARCH, 1945

AN I.B.M. TECHNIQUE FOR THE COMPUTATION OF $\sum X^2$ AND $\sum XY$*

KURT BENJAMIN
GRADUATE RECORD EXAMINATION

Given I.B.M. cards punched with scores (or any numbers)—but not their squares—a method is presented of tabulating them (on the No. 405 alphameric I.B.M. tabulator) so as to obtain the sum of squares. The technique is also adaptable to summation of cross-products. The principle is an extension of the Mendenhall-Warren-Hollerith technique of vertical progressive digiting, without the necessity of manual addition or summary-punching, and is designed for machines not equipped with the "card cycle total transfer" device or "progressive total" device. Use is made of "counter rolling." Efficient use of machine capacity is made only when intercorrelations between no more than two variables are required *in addition to* sums of squares. A resumé of some techniques now commonly employed is included.

To obtain a sum of squares, the following are some of the I.B.M. methods now in common use, when detail cards contain scores (or any numbers), but not their squares:

A. Selecting one master squares card for every detail card. This requires a collator and a previously prepared master file (which may, however, also contain other data—such as higher powers). Matched masters are subsequently tabulated, and then re-merged with the file (8).

B. Interspersed master gang-punching. This also requires prepared squares deck, use of gang-punch machine, and subsequent tabulation of detail cards.

C. Use of automatic multiplying punch. Summary-products counter will contain sum of squares; or, squares can be punched into details and these tabulated (3).

D. Horizontal digiting, requiring at least one digit selector, and three counter-groups per variable. This method also requires multiplication and addition of totals (5).

E. Mendenhall-Warren-Hollerith correlation method: printing of

* The author is indebted to Dr. Paul Dwyer, Associate Professor of Mathematics, University of Michigan, for valuable criticism of the original draft; and to Mr. Alan Meacham, in charge of the University's Tabulating Station, for testing the method.

vertically progressively digited totals—from highest to lowest score. This requires "progressive total" device, manual addition after tabulation and allowance for gaps in the distribution (1, 2, 9).

F. Summary- punching vertically progressively digited totals. This requires digit cards (for possible gaps in the distribution), summary-punch in conjunction with tabulator (equipped with "progressive total" device), and subsequent tabulation of summary cards.

G. Transfer of vertically progressively digited totals to another counter (which will finally contain the sum of squares) by use of "card cycle total transfer" device. Digit cards are required, but no "progressive total" device (6).

Sorting is required in all cases, except (C) and (D).

The method here described is similar to (G), but designed for machines not equipped with "card cycle total transfer" device. It is based on the same mathematical principle as (E), (F), and (G) (see 4). A detailed explanation of the wiring, with explanatory notes, is given in an Appendix.

Necessary Equipment

a) Sorter; b) alphameric tabulator No. 405 (I.B.M.), equipped with at least 2 class selectors, and a number of independent "X"-distributors equal to the number of digits expected in the sum of scores (5 were allowed for in the appended wiring directions). "D" pick-up hubs must be operative; but neither "progressive total" device nor digit selector is needed. c) 2 digit cards for every possible score from the highest actual down to 1 (there should be none for zero scores), with scores punched in the same fields used in detail cards. Digit cards should contain an identifying punch, and be sorted *behind* details, all in order from highest to lowest score.

For example, assume the following scores:

STUDENT NO:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
SCORE:	12	4	12	9	11	9	11	8	11	11	0

The order of cards will be:

Detail	card with	12	in columns	7-8		
"	" "	12	" "	7-8		
Digit	" "	12	" "	7-8	and "X" in column	80
"	" "	12	" "	7-8	" " " "	80
Detail	" "	11	" "	7-8		
"	" "	11	" "	7-8		
"	" "	11	" "	7-8		

Detail	card	with	11	in	columns	7-8	and	"X"	in	column	80
Digit	“	“	11	“	“	7-8	“	“	“	“	80
“	“	“	11	“	“	7-8	“	“	“	“	80
“	“	“	10	“	“	7-8	“	“	“	“	80
“	“	“	10	“	“	7-8	“	“	“	“	80
Detail	“	“	09	“	“	7-8					
“	“	“	09	“	“	7-8					
Digit	“	“	09	“	“	7-8	“	“	“	“	80
“	“	“	09	“	“	7-8	“	“	“	“	80
Detail	“	“	08	“	“	7-8					
Digit	“	“	08	“	“	7-8	“	“	“	“	80
“	“	“	08	“	“	7-8	“	“	“	“	80
“	“	“	07	“	“	7-8	“	“	“	“	80
“	“	“	07	“	“	7-8	“	“	“	“	80
“	“	“	06	“	“	7-8	“	“	“	“	80
“	“	“	06	“	“	7-8	“	“	“	“	80
“	“	“	05	“	“	7-8	“	“	“	“	80
“	“	“	05	“	“	7-8	“	“	“	“	80
Detail	“	“	04	“	“	7-8					
Digit	“	“	04	“	“	7-8	“	“	“	“	80
“	“	“	04	“	“	7-8	“	“	“	“	80
“	“	“	03	“	“	7-8	“	“	“	“	80
“	“	“	03	“	“	7-8	“	“	“	“	80
“	“	“	02	“	“	7-8	“	“	“	“	80
“	“	“	02	“	“	7-8	“	“	“	“	80
“	“	“	01	“	“	7-8	“	“	“	“	80
“	“	“	01	“	“	7-8	“	“	“	“	80
Detail	“	“	00	“	“	7-8					

Machine Principle

After all detail cards containing a particular score have passed the add brushes, the total of scores accumulated in a counter group (hereafter referred to as the transmitting counter) is transferred to another counter group (henceforth referred to as the receiving counter), while the *first* digit card of that score group passes the lower brushes. As the *second* digit card passes the add brushes, the figure in the transmitting counter is restored to the total it contained prior to transfer. Thus, the receiving counter will finally contain a sum of the progressive totals, i.e., the sum of squares (digit cards take care of possible gaps in the distribution). The transmitting counter will print the final progressive total, i.e., the sum of scores.

It is obvious that this technique is easily adapted to summation of cross-products by transferring the progressive totals of one variable at each change in the score of the other variable.

A permanent file of 18 (twice 9) digit cards, each punched with its particular digit in all card fields, would obviate digit card preparation for each job, if control is on *one* column at a time. This method can be used with a multiple digit score if the whole score is added each time—partial products method. It is then necessary to multiply the totals of the 10's position tabulation by 10, of the 100's position by 100, etc., and add them to the units position total.

Limitations

In view of the large number of selectors required and the relatively inefficient utilization of counter capacity, this method is recommended only when sums of squares, but no intercorrelations, are needed, or when the latter are confined to two variables.

REFERENCES

1. Brandt, A. E. The use of machine factoring in multiple correlation. *J. Amer. Stat. Assn.*, 1928, 23, 291-295.
2. Brandt, A. E. Uses of the progressive digit method. Practical applications of the punched card method in colleges and universities. New York: Columbia University Press, 1935. Edited by G. W. Baehne.
3. Carver, H. C. Uses of the automatic multiplying punch. Practical applications of the punched card method in colleges and universities. New York: Columbia University Press, 1935. Edited by G. W. Baehne.
4. Dwyer, Paul S. The computation of moments with the use of cumulative totals. *Ann. math. Statistics*, 1938, 9, 288-304.
5. Dwyer, Paul S. Summary of problems in the computation of statistical constants with sorting and tabulating machines. Proceedings of the Educational Research Forum, Endicott, N. Y., August, 1940 (I.B.M.) See especially p. 24, (Milliman method and I.B.M. improvement).
6. International Business Machines Corp., Card cycle total transfer. Commercial Research Bulletin No. 60, March 5, 1941.
7. Lorge, Irving. Tabulating and test-scoring machines: applications of International Business Machines to educational research. *Rev. educ. Res.*, 1942, 12, 550-557.
8. Meacham, Alan D. The value of the collator in using prepunched cards for obtaining moments and product moments. *Proceedings of the Educational Research Forum*, Endicott, N. Y., August, 1940 (I.B.M.)
9. Mendenhall, R. M., and Warren, Richard. The Mendenhall-Warren-Hollerith Correlation Method. Columbia University Statistical Bureau Document No. 1, 1929. Columbia University, New York.

APPENDIX

WIRING DIRECTIONS: (Assume two-digit scores in columns 7-8, and an "X" in column 80 of digit cards. Class selector hubs have been assigned consecutive numbers 1-10, reading from left to right):

From control brush 80 to "X" pick-up of class selector D.
From "X" pick-up of selector D to common position 1 of selector D.
From controlled position 1 of selector D to common position 10 of selector C.
From normal position 10 of selector C to "X" pick-up of selector C.
From add brushes 7-8 to normal positions 9-10 of selector D.
From common positions 5-10 of selector D to entry hubs of counter 6B(19-24).
From common positions 1-6 of selector C to controlled positions 5-10 of selector D.
From "subtract units position control" to normal positions 1-6 of selector C.
From controlled positions of X-distributors 1-5 to controlled positions 1-5 of selector C.
From card count to common position of X-distributors 1-5.
From unequal impulse outlets 1-5 (comparing unit) to "D" pick-ups of X-distributors 1-5.
From "hot nine" hub of any counter(s) (hub to left of S.U.P. entry) to upper comparing relays 1-5.
From "hot nine" hub of counter 6B to S.U.P. entry of counter 6B.
From "hot nine" hub of counter 6D to S.U.P. entry of counter 6D.
From "CI" hub of counter 6B to "C" hub of counter 6B.
From "CI" hub of counter 6D to "C" hub of counter 6D.
From counter total exit 6B(19-24) to counter entry 6D(59-64) and to lower typebars (nine's complement of ΣX).
From counter total exit 6B (positions *20-24* only) to lower comparing relays 1-5.
From counter total exit 6D(59-64) to lower typebars (ΣX^2).
From "plug to 'C' " impulse hub to common position 9 of selector C.
From controlled position 9 of selector C to —(minus) of counter 6B.
From normal position 9 of selector C to common position 2 of selector D.
From normal position 2 of selector D to —(minus) of counter 6B.
From controlled position 2 of selector D to +(plus) of counter 6B.
From +(plus) of counter 6B to —(minus) of counter 6D.
From "minor class of total" to "class of total control" of counter 6B.
From "intermediate class of total" to "class of tot. contr." of 6D.

(The two counter groups must be reset on different total *cycles*. Otherwise the balance test and selector "total" impulses preceding total cycles will prevent the left-most position of 6D from resetting, whenever the equivalent position of counter 6B contains a 9—as it will, in view of the use of complements). Automatic checks for the presence and correct position of digit cards can easily be devised.

EXPLANATORY NOTES: It will have been observed in the above wiring directions that, in order to transfer the total from one counter to another, 10 (S.U.P.C. = 10) was added to the number in each position of the transmitting counter, thus causing 1's to carry-over. The second digit card allows time for correction of the carry-over, thus restoring the correct total. However, whenever a 9 stands in a counter position and a 10 is added in that position and the position to its right, the expected carry-over of 2 into the position to the left of the first-mentioned does not occur, since the machine will carry only 1 on any one

cycle. This left position must therefore *not* be corrected for carry-over. This is the reason for the comparing unit plugging; e.g.:

hypothetical counter total	0	3	9	6	before transfer.
first digit card	10	10	10	10	(S.U.P.C. for transfer).
counter total after transfer:	1	4	0	6	*NOT 1506*
second digit card:	1	-	1	-	(—1 after transfer)
after correction	0	3	9	6	= as before transfer.

If a "1" had also been deducted in the 100's position, an incorrect total would have been obtained (units position is never corrected, since 100,000's position, from which it receives its carry-over—"C.I." to "C"— should always be at 9: *vide* below for reason for use of complements).

Since it is necessary that the transmitting counter add, and the receiving counter subtract, during transfer, the scores also were subtracted into the transmitting counter, so that the sum of squares would be a true figure (complement of a complement). The sum of scores will be a nine's complement, unless printed from an independently cumulating counter group. Use of ten's complements entails sacrifice of one receiving counter position.

NOTE: Progressive totals can be "counter-listed" from receiving counter during transfer in *alphameric* typebars (numeric typebars would print symbol whenever a 9 is being transferred).

TOTALS STANDING IN BOTH COUNTER-GROUPS, **after each card has passed the add brushes, assuming hypothetical scores given in body of article:**

CARD NO.	*COLS. 7-8*	*COL. 80*	*COUNTER 6 B (transmitting)*	*COUNTER 6 D (receiving)*
1st	12	—	999987	000000
2nd	12	—	999975	000000
3rd	12	X	000086	000024
4th	12	X	999975	000024
5th	11	—	999964	000024
6th	11	—	999953	000024
7th	11	—	999942	000024
8th	11	—	999931	000024
9th	11	X	000042	000092
10th	11	X	999931	000092
11th	10	X	000042	000160
12th	10	X	999931	000160
13th	09	—	999922	000160
14th	09	—	999913	000160
15th	09	X	000024	000246
16th	09	X	999913	000246
17th	08	—	999905	000246
18th	08	X	000016	000340
19th	08	X	999905	000340
20th	07	X	000016	000434
21st	07	X	999905	000434
22nd	06	X	000016	000528
23rd	06	X	999905	000528
24th	05	X	000016	000622
25th	05	X	999905	000622
26th	04	—	999901	000622
27th	04	X	000012	000720
28th	04	X	999901	000720
29th	03	X	000012	000818
30th	03	X	999901	000818
31st	02	X	000012	000916
32nd	02	X	999901	000916
33rd	01	X	000012	001014
34th	01	X	999901	001014
35th	00	—	999901	001014
	Nine's complement of ΣX:		999901	ΣX^2: 1014
		ΣX:	98	

THURSTONE, L. L. *A Factorial Study of Perception.* Chicago: University of Chicago Press, 1944. Pp. vi + 148.

A REVIEW

This imposing study sets forth the Pearson product-moment correlations among 60 selected variables, and reports a multiple-factor analysis of 43 of these variables. The 17 scores excluded from the factor analysis were judged free from significant common-factor variance by reason of their insignificant correlations with other variables in the battery. Most of the tests retained for the factor analysis are tests of visual perception; some of the tests are new. In selecting the perceptual measures, preference was given "to those perceptual effects which . . . might conceivably be centrally [rather than peripherally] determined" (p. 1). The subjects for the study consisted mostly of University of Chicago undergraduate volunteers, the typical correlation being based on about 170 cases. The tests were given in four sessions—three sessions for the individual laboratory tests, and one for the group tests.

Among the perceptual tests employed were such as the following: Street gestalt completion; peripheral visual span; hidden digits; autokinetic movement; flicker fusion; spiral aftermovement; Müller-Lyer illusion; size-weight illusion; Schmidt color-form ratio; color-form sorting; retinal rivalry reversals; shape constancy; and hidden pictures. The battery, however, was not limited to perceptual tests alone. Tests which the present reviewer would classify as partly perceptual and partly intellectual are the Gottschaldt figures, the Kohs Block Designs, and the Space tests of the Primary Mental Abilities battery. Purely intellectual variables included scores V, N, W, and R from the Primary Mental Abilities battery. In addition, the battery of 60 variables included tests of which the following may be selected as representative: reaction time to sound; speed of dark adaptation; speed of judgment; visual-motor coordination (two-hand); free association to selected verbal stimuli; and a modification of mirror drawing.

Specifically excluded from the main study were all direct measurements of personality (such as the Guilford personality schedule). It is the author's avowed goal, eventually, to obtain measures of personality through the perceptual tests. Two major advantages of such perceptual measures of personality would be objectivity and freedom from deception—both deliberate deception and unconscious self-deception. Underlying the hope that objective perceptual measures may yield measures of personality is the broad theory of interdependence of all functions of the individual. According to Thurstone, "it would be difficult to maintain that any of these functions, such as perception, is isolated from the rest of the dynamical system that constitutes the person" (p. 3). In passing, the reviewer must confess that he finds no difficulty at all in maintaining that many independent or "isolated" functions do exist. For example: Visual or tactual perception is basic to (say) reading, and reading is practically essential both to learning and the attitudes that learning can promote. But vision may be excellent and reading quite copious—yet learning may be very poor, and attitudes remain nearly at the level of the animal kingdom. The evidence from correla-

tional studies suggests that the philosophical holism of the gestaltists is a dogma, and the "integration" of the mental hygienists only a hope or ideal. Indirect evaluation of personality—whether by such measures as perceptual tests, the interpretation of Rorschach scores, or whatnot—may be possible; but the burden of proof falls heavily on the advocates of such indirect measures. Thurstone appears to appreciate this when he remarks, "Such an approach as here represented is very much of a gamble" (p. 2).

The factor analysis of the 43 selected variables was carried out by Thurstone's multiple-factor technique; the rotational problem was handled by plotting the graphs for paired columns of the factor-matrix on the printing tabulating machine, using a method devised by Ledyard Tucker. Most of the correlations among the final factors are near zero; none exceeds .25 (p. 123). Very briefly, the 11 factors extracted were interpreted by Thurstone as representing:

A. Facility and firmness in perceptual closure (shape constancy, Gottschaldt figures, Spatial score (from Primary Mental Ability battery)).

B. Error in response to optical illusions (Müller-Lyer, Poggendorf, etc.).

C. Reaction time (light and sound).

D. Perceptual oscillation (retinal rivalry, Necker cube).

E. Freedom from *Gestaltbindung*, flexibility in manipulating several more or less irrelevant or conflicting configurations simultaneously or in succession (two-hand coordination, hidden pictures, Reasoning (from Primary Mental Abilities battery).

F. Speed of perception (perceptual span, dark adaptation, Gottschaldt figures, Street gestalt completion).

G. A second-order general factor common to the composite scores on the Primary Mental Abilities.

H. A factor reflecting performance on the Schmidt apparatus for differentiating between form- and color-dominance.

J. Speed of judgment (color-form sorting time; social judgments time).

K. Rorschach Test scores (total number of responses; and tendency toward "whole" responses).

L. An unidentified residual factor.

It is obvious that some of these factors represent an addition to our understanding of the perceptual field. The reviewer would like to qualify Factor F to the extent of suggesting that this factor seems, for the most part, to reflect speed in perception based on minimal cues. With regard to the findings on speed, Thurstone remarks that the results "point to the desirability of investigating speed in distinguishable functions, rather than speed as a generalized factor" (p. 119). It appears of some special interest that ". . . the well-known optical illusions do have a factor in common, in spite of the fact that their intercorrelations are on the whole rather low" (p. 120). Although the Primary Mental Abilities are in general unrelated to the perceptual factors, the Space factor is definitely related to Factor A (loading of .54) and the Reasoning factor to Factor F (loading of .55).

An important finding relates to color-form dominance. The two tests which appear to be the best measure of this function are (*a*) the Schmidt color-form ratio, and (*b*) color-form sorting (variables 16 and 34, respectively). The product-moment correlation between these two variables is −.013. As Thurstone puts it, "Color and form dominance did not emerge as a distinguishable category" (p. iv). The suggestion is made that surface texture may be a quality requiring special consideration in the study of the color-form problem (pp. 116, 122).

The Rorschach Test does not draw favorable comment in this study. Probably those inclined to defend the Rorschach would say that Rorschach "scores" do not belong in a correlational analysis. The complete configuration of scores must first be submitted to "interpretation" according to Rorschach principles. These *interpretations* may then be included in the list of variables (provided, of course, that the interpretations are quantitatively stated); but the original Rorschach scores taken alone are virtually meaningless.

The last chapter of the monograph reports results from the application of perceptual and other tests to four special groups: a University of Chicago group of fast *vs.* slow readers; a University of Chicago group of campus leaders; a Washington, D.C., group of interns in public administration, who had been rated on professional promise and success; a group of Washington, D.C., administrators classified as (*a*) analysts vs. (*b*) personnel men; and finally, the same group of adminstrators classified according to salary (adjusted for age). Space is lacking for comment on these studies, except to say that some reliable test-differentiations were achieved, and leads for further investigation noted. These studies of special groups were exploratory and are only very briefly reported. In the study of these special groups, the statistical procedure consisted of preparing four-fold tables and determining chi-square. It goes hard with this reviewer to see quantitative data, such as test-scores or salaries, reduced to two categories only.

One statistical comment applies to the whole monograph: not a single reliability coefficient is reported, either for test-scores or factor-scores. This appears rather undesirable even in a "frankly exploratory" study (p. 101), particularly since some of the tests are new, or have been modified in presentation or scoring.

A study on a scale so large as this one involves a heavy burden of apparatus-construction, test administration, test scoring, statistical analysis, and professional cooperation. Thurstone's indebtedness to his colleagues and assistants is appropriately recorded both in the Preface and throughout the text.

In summary, this monograph makes a significant contribution both in the presentation of new tests, and in the study of interrelations among selected variables centering about perception. New understanding has been gained of the perceptual field. In addition, a series of studies has been executed, exploring the value of perceptual and other tests in the differentiation of important special groups. These contributions are more than sufficient to make this monograph a landmark of current psychological progress.

Herbert S. Conrad
College Entrance Examination Board
Princeton, N. J.